

CP	1024	CAGACGGAGTTGAGATGCTGGAGGAGATGAGATGAGTCACTCTCTAACTTAACAGACCCACGACG	965
Db	244	ACTTATCCACGCGGAGAGCTCTCCGCCACCCGAGTCTCTCCCATTTCTTCTACTT	303
CP	964	ACTTATCCACGCGGAGAGCTCTCCGCCACCCGAGTCTCTCCCATTTCTTCTACTT	905
Db	304	TGCGCAGTTCAGAGTGTCTCTCTCCACCACTCCGACAAACCTCAATTAATCAAGAG	363
CP	904	TGCGCAGTTCAGAGTGTCTCTCTCTCCACCACTCCGACAAACCTCAATTAATCAAGAG	845
Db	364	ACCTGATTTA 374	
CP	844	ACCTGATTTA 834	
RESULT	13		
LOCUS	AA047184	355 bp	EST
DEFINITION	ZK74802.s1	Scores pregnant uterus NBHPU Homo sapiens CDNA clone	06-SEP-1996
FEATURES	488522..37		
ASSIGN	AA047184		
NID	91525150		
KEYWORDS	EST		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote: mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 355)		
AUTHORS	Hallier, L., Clark, N., Dubouq, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Woldmann, P., and Wilson, R.		
TITLE	The Mashu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK Mashu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.lln.gov) for further information. Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 374. Location/Qualifiers 1. 355 /organism="Homo sapiens" /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGAATTCGCGGCGCCCTTTTCTTTTCTTTTCTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo." /db_xref="taxon:9606" /clone="488522" /clone_1id="Scores pregnant uterus NBHPU" /sex="female" /dev_stage="adult" /lab_host="DH10B" complement(<1..>355) BASE COUNT 91 a 121 c 66 g 76 t 1 others ORIGIN		
Query Match	28.6%	Score 351; DB 14; Length 355;	
Best Local Similarity	99.7%	Pred. No. 0.00e+00;	
Matches	351; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
4	GACAGTAATCACTTTTATTGTGTTCACAGAAATACTAGCGAGTCTGACAGTGCCTC	63	

CP	1204	GACACTAAATCAATTATTTTATTTGTGTTACAGAACTACTAGGCGATCTCGAAGTGGCTC	1145			
Db	64	CGTGACACCCACACACACCCCAACCCCTACTACTCCAGCAGCACCTTAAAGGCGACTTCAAG	123			
CP	1144	CGTGACACCCACACACACCCCAACCCCTACTACTCCAGCAGCACCTTAAAGGCGACTTCAAG	1085			
Db	124	AAGATGAAAGATCTCACGGATCTCATTTCTTAATGGTCCGCCGAAGTCTCACACAGTAGA	183			
CP	1084	AAGATGAAAGATCTCACGGATCTCATTTCTTAATGGTCCGCCGAAGTCTCACACAGTAGA	1025			
Db	184	CAGACGGAGTTAGATGCTGGAGGATGAGACGTACCTCCAACTTAACGACCCACACCCAG	243			
CP	1024	CAGACGGAGTTAGATGCTGGAGGATGAGACGTACCTCCAACTTAACGACCCACACCCAG	965			
Db	244	ACTTCATCCACAGCCGAGAGCTCTCCACCCAGAGTCTCCCAATTTTCTTCAGACT	303			
CP	964	ACTTCATCCACAGCCGAGAGCTCTCCACCCAGAGTCTCCCAATTTTCTTCAGACT	905			
Db	304	TGCCGCACTCCAGGTGCTCTCTTCACCAAGTCCCAACCAAGCTCAATTAAT	355			
CP	904	TGCCGCACTCCAGGTGCTCTCTTCACCAAGTCCCAACCAAGCTCAATTAAT	853			
RESULT	14	T49532	405 bp	mRNA	EST	08-FEB-1995
LOCUS	DEFINITION	ya76f12.s1	Homo sapiens	CDNA	clone 67631	37
ACCESSION		T49532				
KEYWORDS		G651392				
SOURCE		EST.				
REFERENCE		human clone-67631 library-Stratagene placenta (#937225)				
AUTHORS		vector-pbluescript SK- host-SOLR cells (kanamycin resistant)				
		primer--2lm13 Ralte-EcORI Ralte-EcORI XhoI placental tissue from a				
		Caucasian male. Cloned unidirectionally. Primer: Oligo dt. Average				
		insert size: 1.2 kb; Uni-ZAP XR Vector; 5' adaptor sequence:				
		5'-GAATTCGGCAGCAG-3'; 3' adaptor sequence:				
		5'-CTCAGAGTTTATTTTATTTT-3'.				
ORGANISM		Homo sapiens				
		Eucaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia				
		Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE		1 (bases 1 to 405)				
AUTHORS		Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,				
		Holman, M., Hultman, M., Kuchaba, T., Le, M., Lennon, G., Marra, M.,				
		Parsons, J., Rifkin, L., Ruhlberg, T., Tan, F., Trevisan, E.,				
		Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.				
TITLE		WashU-Merck EST Project				
JOURNAL		Unpublished (1995)				
COMMENT		Other-ESTs: ya76f12.r1				
		Contact: Wilson RK				
		WashU-Merck EST Project				
		Washington University School of Medicine				
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
		Tel: 314 286 1800				
		Fax: 314 286 1810				
		Email: est@watson.wustl.edu				
		Source: IMAGE Consortium, LNL				
		This clone is available royalty-free through LNL; contact the				
		IMAGE Consortium (info@image.lnl.gov) for further information.				
FEATURES		Location/Qualifiers				
Source		1..405				
		/organism="Homo sapiens"				
		/clone="67631"				
BASE COUNT		108 a 132 c 81 g 82 t 2 others				
ORIGIN						
Query Match		28.1%; Score 345; DB 9; Length 405;				
Best Local Similarity		96.3%; Pred. No. 0.00e+00;				
Matches		388; Conservative 0; Mismatches 9; Indels 6; Gaps 6;				
Db	1	TGTGTTACAGAACTACTAGCGATCTCGACAGTCTCGTGTGACAGCCCAACCC	60			
CP	1184	TGTGTTACAGAACTACTAGCGATCTCGACAGTCTCGTGTGACAGCCCAACCC	1125			

Db 61 CAACCTTACCTGCGACGCAACCCCTAAGGCGACTTCAAGAAGATGGAAGATCTCAGC 120
 |||||||
 Cp 1124 CAACCTTACCTGCGACGCAACCCCTAAGGCGACTTCAAGAAGATGGAAGATCTCAGC 1065
 |||||||
 Db 121 ATCTCATCTTAATGGTCCGCGAAGTCTCACAGTACAGAGAGAGAGTGAATGCTG 180
 |||||||
 Cp 1064 ATCTCATCTTAATGGTCCGCGAAGTCTCACAGTACAGAGAGAGAGTGAATGCTG 1005
 |||||||
 Db 181 GAGATGAGTACCTCTCTTAACCTTACGACCCACAGACAGCTTATCTCCAGCGGAGC 240
 |||||||
 Cp 1004 GAGATGAGTACCTCTCTTAACCTTACGACCCACAGACAGCTTATCTCCAGCGGAGC 945
 |||||||
 Db 241 TCTCTCCCGCCGAGCTCTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
 |||||||
 Cp 944 TCTCTCCCGCCGAGCTCTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 886
 |||||||
 Db 301 CTGCTTCACGAGTCCCAAAAGCTCAATAAATACAGAGAGAGAGCTGATTCAGAGCA 360
 |||||||
 Cp 885 CTGCTTCACGAGTCCCAAAAGCTCAATAAATACAGAGAGAGAGCTGATTCAGAGCA 829
 |||||||
 Db 361 GAGGGAGACATNTNACACCTTGGCATAGGTTTAAATAAT 403
 |||||||
 828 GAGGG-AACATCTCACACCTTG-CATAAGTTAAATAATAAT 788
 |||||||

RESULT 15
 LOCUS AA301628 413 bp mRNA EST 18-APR-1997
 DEFINITION EST14813 Aorta endothelial cells Homo sapiens cDNA 5' end.
 ACCESSION AA301628
 MID g1954112
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homidae;
 Homo.

REFERENCE 1 (bases 1 to 413)
 AUTHORS Adams,M.D., Kerlavage,A.R., Fieschmann,R.D., Fuldner,R.A.,
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagan,N.S.,
 Glodet,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Utechtback,T.R., Weidman,J.F., Li,Y.,
 Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Danke,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A.,
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
 Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
 Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
 Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
 Fraser,C.M., and Venter,J.C.
 TITLE Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 96026280
 COMMENT Other ESTs: TH0175266
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org

For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 1. 413
 Location/Qualifiers
 source
 /organism="Homo sapiens"

/note="Organ: aorta; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI"
 /db_xref="ATCC (inhost):113941"
 /db_xref="taxon:9606"
 /clone_lib="Aorta endothelial cells"
 /cell_type="endothelial cell"
 /dev_stage="adult"
 mRNA
 BASE COUNT 105 a 108 c 99 g 95 t 6 others
 ORIGIN
 Query Match 27.0%; Score 332; DB 18; Length 413;
 Best Local Similarity 96.8%; Pred. No. 0.00e+00;
 Matches 364; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
 Db 1 CTGACCTCAGACCTTCTCTCCAGCAACACCGGCGCTTCTACTGTAATCTCCCGCA 60
 |||||||
 Qy 427 CTGACCTCAGACCTTCTCTCCAGCAACACCGGCGCTTCTACTGTAATCTCCCGCA 486
 |||||||
 Db 61 CTCCCAAAAGATCCGAAACCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 120
 |||||||
 Qy 487 CTCCCAAAAGATCCGAAACCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 546
 |||||||
 Db 121 N-TATCCCACTGGAGCTTCCGAGCAACTTGAACCTCAGAACTACACAGAGAGAGC 179
 |||||||
 Qy 547 CGTATCCCACTGGAGCTTCCGAGCAACTTGAACCTCAGAACTACACAGAGAGAGC 606
 |||||||
 Db 180 ACCCGGTCTTGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 239
 |||||||
 Qy 607 ACCCGGTCTTGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 666
 |||||||
 Db 240 CAGCCAGCTGGGGGCTGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 299
 |||||||
 Qy 667 CAGCCAGCTGGGGGCTGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 725
 |||||||
 Db 300 CTNCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 359
 |||||||
 Qy 726 CTCTT-AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 782
 |||||||
 Db 360 ACGTAATTAATTAATTT 375
 |||||||
 Qy 783 ACGTAATTAATTAATTT 798
 |||||||

Search completed: Thu Apr 23 05:39:21 1998
 Job time : 811 secs.

Query Match 8.8%; Score 98; DB 1; Length 475;
 Best Local Similarity 28.6%; Pred. No. 6.05e-01;
 Matches 18; Conservative 19; Mismatches 24; Indels 2; Gaps 2;

DB 358 LIGLLEFDKVTDMRLAMEE-PRGPVLPILRTVSVEEALISNK-SFYQASIFNDP 415
 QY 94 VFQCLLAEGVPAPLPEDAPNAPASLAPTPVSPLEPFLTSEPSDYALDSTFLOQHP 153
 DB 416 RAF 418
 QY 154 AAF 156

RESULT 15
 ID A1AB_HUMAN STANDARD; PRT; 519 AA.
 AC P35368;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE ALPHA-1B ADRENERGIC RECEPTOR.
 GN ADRA1B.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 EC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93016158.
 RA RAMARAO C.S., DENKER J.M., PEREZ D.M., GALVIN R.J., RIEK R.P.,
 RA GRAHAM R.M.,
 RL J. BIOL. CHEM. 267:21936-21945(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 94239386.
 RA FORRAY C., BARD J.A., WETZEL J.M., CHIU G., SHAPIRO E., TANG R.,
 RA LEPOR H., HARTIG P.R., WEINSHANK R.L., BRANCHER T.A.,
 RA GLUCHOWSKI C.,
 RL MOL. PHARMACOL. 45:703-708(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95114877.
 RA SCHWAIN D.A., JOHNSTON G.I., PAGE S.O., MOSLEY M.J., WILSON K.H.,
 RA WOMAN N.P., CAMPBELL S., FIDOCK M.D., FURNESS L.M.,
 RA PARRY-SMITH D.J., PETER B., BAILEY D.S.,
 RL J. PHARMACOL. EXP. THER. 272:134-142(1995).
 CC -1- FUNCTION: THIS ALPHA-ADRENERGIC RECEPTOR MEDIATES ITS ACTION BY
 ASSOCIATION WITH G PROTEINS THAT ACTIVATE A PHOSPHATIDYLINOSITOL-
 CALCIUM SECOND MESSENGER SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 EMBL: M99589; -; NOT ANNOTATED_CDS.
 EMBL: U03865; G494983; -;
 DR EMBL: L31773; G666891; -;
 DR PIR: A45121; A45121.
 DR GCRDB: GCR_0530; -;
 DR MIM: 104220; -;
 DR PROSITE: PS00237; G-PROTEIN RECEPTOR; 1.
 DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
 KW MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
 FT DOMAIN 1 45
 FT TRANSMEM 46 70
 FT DOMAIN 71 83
 FT TRANSMEM 84 105
 FT DOMAIN 106 115
 FT TRANSMEM 116 141
 FT DOMAIN 142 161
 FT TRANSMEM 162 182
 FT DOMAIN 183 201
 FT TRANSMEM 202 224
 FT DOMAIN 225 295
 FT TRANSMEM 296 319
 FT DOMAIN 320 326
 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 327 340 7 (POTENTIAL).
 FT DOMAIN 341 519 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 370 379 POLY-ARG.
 FT CARBOHYD 10 10-
 FT CARBOHYD 24 24 POTENTIAL.
 FT CARBOHYD 29 29 POTENTIAL.
 FT CARBOHYD 34 34 POTENTIAL.
 FT DISULFID 118 195 BY SIMILARITY.
 FT LIPID 365 365 PALMITATE (POTENTIAL).
 FT CONFLICT 379 379 MISSING (IN REF. 1).
 FT CONFLICT 497 500 AAAD -> PRH (IN REF. 1).
 FT CONFLICT 370 370 R -> RG (IN REF. 2).
 SQ SEQUENCE 519 AA; 56778 MW; D088058C CRC32;

Query Match 8.8%; Score 98; DB 1; Length 519;
 Best Local Similarity 41.3%; Pred. No. 6.05e-01;
 Matches 19; Conservative 9; Mismatches 15; Indels 3; Gaps 3;

DB 451 ALLSLPAPPPGRRG-RHDSGP-LFTFKLLAPES-PGTDGASNG 493
 QY 12 TILQAPTPAPSTIPGPRGSGPEITFTDPLPEPAPAPGRRSASRG 57

Search completed: Fri Apr 17 13:38:42 1998
 Job time : 10 secs.

OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EC16;
RX MEDLINE: 93054355.
RA LINDBERG M., COLLIER A.;
J. BACTERIOL. 174:7385-7397(1992).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
MULTIPLE PEPTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
DR EMBL: L02214; G148433; -.
DR PIR: A47021; A47021.
DR PROSITE: PS01141; T2SP-C; 1.
KW TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
TRANSMEM 17 35 PERIPLASMIC (POTENTIAL).
DOMAIN 36 272 PERIPLASMIC (POTENTIAL).
SEQUENCE 272 AA; 30301 MW; 7944F46D CRC32;
Query Match 8.9%; Score 99; DB 1; Length 272;
Best Local Similarity 36.7%; Pred. No. 4,56e-01;
Matches 22; Conservative 11; Mismatches 24; Indels 3; Gaps 3;
Db 6 LPLSPSVIRILFYLLMLFCCQL-AMIFWVGLP-DNSP-VASVOITPAQARQOPVTL 62
75 LPVEEPNPAKRLFLFLTLTYFCQILMAEGVPAPLPEDAPNNAASLAPTPVSPVLEPFNL 134
RESULT 12
ID GSOC_ERWCH STANDARD; PRT; 272 AA.
AC 001564;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GENERAL SECRETION PATHWAY PROTEIN C (PEPTIC ENZYMES SECRETION PROTEIN
OUTC).
GN OUTC.
OS ERWINIA CHRYSANTHEMI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3937;
RX MEDLINE: 93086427.
CONDEMNINE G., DOREL C., HUGOUIREUX-COTTE-PATTAT N., ROBERT-BAUDOUY J.;
MOL. MICROBIOL. 6:3199-3211(1992).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
MULTIPLE PEPTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
DR EMBL: X65265; G42201; -.
DR PIR: S28013; S28013.
DR PROSITE: PS01141; T2SP-C; 1.
KW TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
TRANSMEM 17 35 PERIPLASMIC (POTENTIAL).
DOMAIN 36 272 PERIPLASMIC (POTENTIAL).
SEQUENCE 272 AA; 30162 MW; 25634B57 CRC32;
Query Match 8.8%; Score 98; DB 1; Length 272;
Best Local Similarity 36.7%; Pred. No. 6,05e-01;
Matches 22; Conservative 11; Mismatches 24; Indels 3; Gaps 3;
Db 6 LPLSPSVIRILFYLLMLFCCQL-AMIFWVGLP-DNAP-VSSVOITPAQARQOPVTL 62
75 LPVEEPNPAKRLFLFLTLTYFCQILMAEGVPAPLPEDAPNNAASLAPTPVSPVLEPFNL 134
RESULT 13
ID VGLI_HSV11 STANDARD; PRT; 390 AA.

AC P06487;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE GLYCOPROTEIN I.
GN GI OR US7.
OS HERPES SIMPLEX VIRUS (TYPE 1 / STRAIN 17).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRINAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85160822.
RA MCGEOCH D.J., DOLAN A., DONALD S., RIXON F.J.;
J. MOL. BIOL. 181:1-13(1985).
CC -1- THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND 2: GH, GB, GC,
GG, GD, GI, AND GE.
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS I, TO VZV GP1V,
AND TO PRV GP63.
DR EMBL: L00036; G291497; -.
DR EMBL: X14112; G58565; -.
DR EMBL: X02138; G59879; -.
DR PIR: A05243; Q0BE77.
KW GLYCOPROTEIN.
FT CARBOHYD 156 156 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT CARBOHYD 257 257 POTENTIAL.
SQ SEQUENCE 390 AA; 41369 MW; 7DA38E2D CRC32;
Query Match 8.8%; Score 98; DB 1; Length 390;
Best Local Similarity 28.4%; Pred. No. 6,05e-01;
Matches 25; Conservative 21; Mismatches 37; Indels 5; Gaps 4;
Db 206 QASTPSTTSTPSTTIPASTTIPAPQASTP-FPTGDKPPQPGVNHPPSNA-TRATR 263
3 HRSCHPTTILQAPTPAST-IPGRGRSGEITFDLPAPAPAGPSPASGHRKR 61
Db 264 DSR-YALTVTQIQIAPASTIALVEL 289
62 SRRVLPVRRVRLQPLVEEPNPAKRLFL 89
RESULT 14
ID GAPN_STRMU STANDARD; PRT; 475 AA.
AC 059931;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.9)
DE (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE)
DE (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (NADP+)) (TRIOSEPHOSPHATE
DEHYDROGENASE).
GN GAPN.
OS STREPTOCOCCUS MUTANS.
OC PROKARYOTA; FIRMICUTES; COCCI; STREPTOCOCCAEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NG5 SEROTYPE C;
RX MEDLINE: 95270576.
BOYD D.A., CVITKOVITCH D.G., HAMILTON I.R.;
J. BACTERIOL. 177:2622-2627(1995).
CC -1- FUNCTION: IMPORTANT AS A MEANS OF GENERATING NADPH FOR
BIOSYNTHETIC REACTIONS.
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + NADP(+) + H(2)O
= 3-PHOSPHO-D-GLYCERATE + NADPH.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: L18521; G642667; -.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; FALSE_NEG.
KW OXIDOREDUCTASE; NADP.
FT NP_BIND 230 235 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 250 250 BY SIMILARITY.
FT ACT_SITE 284 284 BY SIMILARITY.
SQ SEQUENCE 475 AA; 51221 MW; A621BF01 CRC32;

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
 RA SIMS M., SMALLON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLDMAN P.,
 RL NATURE 368:32-38(1994).
 DR EMBL: L16621; G289783; .
 DR PIR: S44920; S44920.
 DR WORMEP: ZK688.5; CE00463.
 DR PROSITE: P550053; UBIQUITIN_2; UNKNOWN_1.
 KW HYPOTHETICAL PROTEIN.
 .. SEQUENCE 1799 AA; 202641 MW; 3EF7DDB2 CRC32;

Query Match 9.8%; Score 109; DB 1; Length 1799;
 Best Local Similarity 23.7%; Pred. No. 2.43e-02;
 Matches 31; Conservative 30; Mismatches 65; Indels 5; Gaps 5;

DB 844 FLENRGRIPSTSPAPSTSENPSPSPNS-EDADIRAGRLPLGTRPNRTVRETVPAA 902
 QY 13 ILQAPRPAPSTIGPRGSGPGLFTDPLPEPAAPAGR-PSASRGKRSKRVLYPRV 71
 DB 903 ARAESPNIHLTFTATTHFPAPAGF-PLMASSNPSTSGAPGWPRIQVSPPTTGL 961
 QY 72 RR-QLPEEPNPAKRLFLTLTIVFCQILMAEGVPAPLP-PEDAPNAASLAPTPVSPVL 129
 DB 962 FEFDLSSSDQ 972
 QY 130 EPNLTSEPSD 140

RESULT 6 STANDARD; PRT; 1206 AA.
 ID FOR4.MOUSE
 AC 005859;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE FORMIN 4 (LIMB DEFORMITY PROTEIN).
 GN LD.
 OS MUS MUSCULUS (MOUSE).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OS EUTHERIA; RODENTIA.
 [1]
 .. SEQUENCE FROM N.A.
 RC TISSUE-EMBRYO;
 RX MEDLINE; 92112033.
 RL GRUSBY-JACKSON L., KUO A., LEDER P.;
 RL GENES DEV. 6:29-37(1992).
 CC -1- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB AND MAY HAVE A
 CC FUNCTION IN DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING
 CC SPECIFIC DIFFERENTIATED STATES.
 CC -1- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT
 CC HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.
 CC -1- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL
 CC ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING
 CC LIMB BUD.
 CC -1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE
 CC PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN
 CC SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE
 CC TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS
 CC DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR
 CC DETERMINED.
 DR EMBL: X62379; G51553; .
 DR PIR: S24407; S24407.
 DR HSSP: P19999; 1CLG.
 KW NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
 FT DOMAIN 635 638 POLY-SER.
 FT DOMAIN 644 744 PRO-RICH.

FT DOMAIN 751 755 POLY-SER.
 SQ SEQUENCE 1206 AA; 133464 MW; 6D70C261 CRC32;

Query Match 9.2%; Score 103; DB 1; Length 1206;
 Best Local Similarity 39.1%; Pred. No. 1.44e-01;
 Matches 18; Conservative 15; Mismatches 9; Indels 4; Gaps 4;

DB 716 VLALNSGGPP-PPPPPPPPGLAPP-PPGL-SFGSSSSSSQRP 758
 QY 98 IL-MAEGVPAPLPEDAPNAASLAPTPVSPVLEPFLNLTSEPSDYA 142

RESULT 7 STANDARD; PRT; 1468 AA.
 ID FOR4.MOUSE
 AC 005860;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE FORMIN (LIMB DEFORMITY PROTEIN).
 GN FMN OR LD.
 OS MUS MUSCULUS (MOUSE).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OS EUTHERIA; RODENTIA.
 [1]
 .. SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY AND TESTIS;
 RX MEDLINE; 90363291.
 RA WOYCHIK R.P., MAAS R.L., ZELLER R., VOGT T.F., LEDER P.;
 RL NATURE 346:850-853(1990).
 CC -1- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
 CC AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY BE
 CC INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
 CC LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
 CC GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.
 CC -1- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN
 CC IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
 CC COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
 CC KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
 CC THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
 CC -1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN
 CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A
 CC VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND
 CC DIFFERENT SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
 CC EMBL: X53599; G52878; .
 DR PIR: S11515; S11515.
 DR HSSP: P19999; 1CLG.
 DR MGD: MGT:101815; FMN.
 KW NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
 FT DOMAIN 298 301 POLY-SER.
 FT DOMAIN 861 864 POLY-SER.
 FT DOMAIN 869 970 PRO-RICH.
 FT DOMAIN 977 981 POLY-SER.
 SQ SEQUENCE 1468 AA; 163809 MW; 42CA4104 CRC32;

Query Match 9.2%; Score 103; DB 1; Length 1468;
 Best Local Similarity 39.1%; Pred. No. 1.44e-01;
 Matches 18; Conservative 15; Mismatches 9; Indels 4; Gaps 4;

DB 942 VLALNSGGPP-PPPPPPPPGLAPP-PPGL-SFGSSSSSSQRP 964
 QY 98 IL-MAEGVPAPLPEDAPNAASLAPTPVSPVLEPFLNLTSEPSDYA 142

RESULT 8 STANDARD; PRT; 268 AA.
 ID CEBD.RAT
 AC 003484;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE CCAT/ENHANCER BINDING PROTEIN DELTA (C/EBP DELTA) (TRANSCRIPTION
 DE FACTOR CELF).

 WISE

 (TM)

Release 3.0.4A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1997, University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm
 on: Fri Apr 17 13:38:32 1998; MasPar time 6.24 Seconds
 Tabular output not generated. 626,917 Million cell updates/sec

Title: >US-08-799-910-10
 Description: (1-156) from US08799910.pep
 Perfect Score: 1114
 Sequence: 1 MCHSRSCHPMTIIOAPPTPA.....EPSDYALDSTFIQHPAAE 156

Scoring table:
 PAM 150
 Gap 11

Searched: 69112 seqs, 25083644 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot35
 1:swiss1

Statistics: Mean 44.412; Variance 91.621; scale 0.485

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Alt	Score	Query	Match	Length	ID	Description	Pred. No.
1	1102	98.9	156	1	IEX1_HUMAN	RADIATION-INDUCIBLE IM	5.97e-191
2	752	67.5	153	1	IEX1_MOUSE	RADIATION-INDUCIBLE IM	1.15e-120
3	117	10.5	474	1	VTP3_TVIV	VIRAL PROTEIN TPX.	2.07e-03
4	112	10.1	228	1	VCOM_ADEM1	MINOR CORE PROTEIN (PR	9.76e-03
5	109	9.8	1799	1	YC25_CAEEL	HYPOHETICAL 202.6 KD	2.43e-02
6	103	9.2	1206	1	FOR4_MOUSE	FORMIN 4 (LIMB DEFORMI	1.44e-01
7	103	9.2	1468	1	FOR4_MOUSE	FORMIN 4 (LIMB DEFORMI	1.44e-01
8	101	9.1	268	1	CEBD_RAT	CCAAT/ENHANCER BINDING	2.57e-01
9	101	9.1	417	1	ALAB_CANFA	ALPHA-1B ADRENERGIC RE	2.57e-01
10	101	9.1	816	1	OAIF_NEUCR	OUTINIC ACID UTILIZATIO	2.57e-01
11	99	8.9	272	1	GSRC_ERWCH	GENERAL SECRETION PATH	4.56e-01
12	98	8.8	272	1	GSRC_ERWCH	GENERAL SECRETION PATH	4.56e-01
13	98	8.8	390	1	VGLT_HSV11	GLYCOPROTEIN 1.	6.05e-01
14	98	8.8	475	1	GAPN_STRMU	NADP-DEPENDENT GLYCERA	6.05e-01
15	98	8.8	519	1	ALAB_HUMAN	ALPHA-1B ADRENERGIC RE	6.05e-01
16	98	8.8	975	1	CDP_CANFA	CCAAT DISPLACEMENT PRO	6.05e-01
17	97	8.7	1043	1	CHS2_PABBR	CHITIN SYNTHASE 2 (EC	8.02e-01
18	97	8.7	1239	1	V120_EBV	CAPSID ASSEMBLY PROTEI	8.02e-01
19	97	8.7	1337	1	PTP4_HUMAN	PROTEIN-TYROSINE PHOSP	8.02e-01
20	96	8.6	245	1	VE4_HPV5B	PROBABLE E4 PROTEIN.	1.06e+00
21	96	8.6	245	1	VE4_HPV5B	PROBABLE E4 PROTEIN.	1.06e+00
22	96	8.6	1233	1	NME3_HUMAN	GLUTAMATE (NMDA) RECEP	1.06e+00
23	95	8.5	234	1	GLNA_DUNSA	GLUTAMINE SYNTHETASE (1.40e+00

24	95	8.5	316	1	CDNC_HUMAN	CYCLIN-DEPENDENT KINAS	1.40e+00
25	95	8.5	449	1 <td>ARG_BRANA <td>ANTER-SPECIFIC PROLINE <th>1.40e+00</th> </td></td>	ARG_BRANA <td>ANTER-SPECIFIC PROLINE <th>1.40e+00</th> </td>	ANTER-SPECIFIC PROLINE <th>1.40e+00</th>	1.40e+00
26	95	8.5	493	1 <td>GUNA_XANCP <td>MAJOR EXTRACELLULAR EN <th>1.40e+00</th> </td></td>	GUNA_XANCP <td>MAJOR EXTRACELLULAR EN <th>1.40e+00</th> </td>	MAJOR EXTRACELLULAR EN <th>1.40e+00</th>	1.40e+00
27	95	8.5	534	1 <td>APG_ARATH <td>ANTER-SPECIFIC PROLINE <th>1.40e+00</th> </td></td>	APG_ARATH <td>ANTER-SPECIFIC PROLINE <th>1.40e+00</th> </td>	ANTER-SPECIFIC PROLINE <th>1.40e+00</th>	1.40e+00
28	95	8.5	643	1 <td>VP40_HSV2 <td>CAPSID PROTEIN P40 (CO <th>1.40e+00</th> </td></td>	VP40_HSV2 <td>CAPSID PROTEIN P40 (CO <th>1.40e+00</th> </td>	CAPSID PROTEIN P40 (CO <th>1.40e+00</th>	1.40e+00
29	95	8.5	3866	1 <td>HRX_MOUSE <td>ZINC FINGER PROTEIN HR <th>1.40e+00</th> </td></td>	HRX_MOUSE <td>ZINC FINGER PROTEIN HR <th>1.40e+00</th> </td>	ZINC FINGER PROTEIN HR <th>1.40e+00</th>	1.40e+00
30	94	8.4	217	1 <td>YKR4_EBV <td>HYPOHETICAL BKRF4 PRO <th>1.84e+00</th> </td></td>	YKR4_EBV <td>HYPOHETICAL BKRF4 PRO <th>1.84e+00</th> </td>	HYPOHETICAL BKRF4 PRO <th>1.84e+00</th>	1.84e+00
31	94	8.4	279	1 <td>Y091_NPVOP <td>HYPOHETICAL 29.3 KD P <th>1.84e+00</th> </td></td>	Y091_NPVOP <td>HYPOHETICAL 29.3 KD P <th>1.84e+00</th> </td>	HYPOHETICAL 29.3 KD P <th>1.84e+00</th>	1.84e+00
32	94	8.4	389	1 <td>NDP_MOUSE <td>NPC DERIVED PROLINE RI <th>1.84e+00</th> </td></td>	NDP_MOUSE <td>NPC DERIVED PROLINE RI <th>1.84e+00</th> </td>	NPC DERIVED PROLINE RI <th>1.84e+00</th>	1.84e+00
33	94	8.4	1742	1 <td>GUNA_CALSA <td>ENOGLUCANASE A PRECUR <th>1.84e+00</th> </td></td>	GUNA_CALSA <td>ENOGLUCANASE A PRECUR <th>1.84e+00</th> </td>	ENOGLUCANASE A PRECUR <th>1.84e+00</th>	1.84e+00
34	93	8.3	232	1 <td>SAX1_CHICK <td>HOMEBOX PROTEIN SAX-1 <th>2.42e+00</th> </td></td>	SAX1_CHICK <td>HOMEBOX PROTEIN SAX-1 <th>2.42e+00</th> </td>	HOMEBOX PROTEIN SAX-1 <th>2.42e+00</th>	2.42e+00
35	93	8.3	268	1 <td>CEBD_MOUSE <td>CCAAT/ENHANCER BINDING <th>2.42e+00</th> </td></td>	CEBD_MOUSE <td>CCAAT/ENHANCER BINDING <th>2.42e+00</th> </td>	CCAAT/ENHANCER BINDING <th>2.42e+00</th>	2.42e+00
36	93	8.3	276	1 <td>YTA4_YEAST <td>HYPOHETICAL 31.3 KD P <th>2.42e+00</th> </td></td>	YTA4_YEAST <td>HYPOHETICAL 31.3 KD P <th>2.42e+00</th> </td>	HYPOHETICAL 31.3 KD P <th>2.42e+00</th>	2.42e+00
37	93	8.3	350	1 <td>Y08N_MYCTU <td>HYPOHETICAL 37.0 KD P <th>2.42e+00</th> </td></td>	Y08N_MYCTU <td>HYPOHETICAL 37.0 KD P <th>2.42e+00</th> </td>	HYPOHETICAL 37.0 KD P <th>2.42e+00</th>	2.42e+00
38	93	8.3	464	1 <td>SP62_HUMAN <td>SPLICOSOME ASSOCIATED <th>2.42e+00</th> </td></td>	SP62_HUMAN <td>SPLICOSOME ASSOCIATED <th>2.42e+00</th> </td>	SPLICOSOME ASSOCIATED <th>2.42e+00</th>	2.42e+00
39	93	8.3	517	1 <td>MBEA_ECOLI <td>MOBILIZATION PROTEIN M <th>2.42e+00</th> </td></td>	MBEA_ECOLI <td>MOBILIZATION PROTEIN M <th>2.42e+00</th> </td>	MOBILIZATION PROTEIN M <th>2.42e+00</th>	2.42e+00
40	92	8.3	620	1 <td>EXTN_TOBAC <td>EXTENSIN PRECURSOR (CE <th>3.17e+00</th> </td></td>	EXTN_TOBAC <td>EXTENSIN PRECURSOR (CE <th>3.17e+00</th> </td>	EXTENSIN PRECURSOR (CE <th>3.17e+00</th>	3.17e+00
41	92	8.3	631	1 <td>DMK_MOUSE <td>MYOTONIN-PROTEIN KINAS <th>3.17e+00</th> </td></td>	DMK_MOUSE <td>MYOTONIN-PROTEIN KINAS <th>3.17e+00</th> </td>	MYOTONIN-PROTEIN KINAS <th>3.17e+00</th>	3.17e+00
42	92	8.3	820	1 <td>FTBL_HUMAN <td>FIBRINOGEN-LIKE PROTEI <th>3.17e+00</th> </td></td>	FTBL_HUMAN <td>FIBRINOGEN-LIKE PROTEI <th>3.17e+00</th> </td>	FIBRINOGEN-LIKE PROTEI <th>3.17e+00</th>	3.17e+00
43	92	8.3	926	1 <td>YTK9_YEAST <td>HYPOHETICAL 103.6 KD <th>3.17e+00</th> </td></td>	YTK9_YEAST <td>HYPOHETICAL 103.6 KD <th>3.17e+00</th> </td>	HYPOHETICAL 103.6 KD <th>3.17e+00</th>	3.17e+00
44	92	8.3	1692	1 <td>POLN_HEVPA <td>NON-STRUCTURAL POLYPRO <th>3.17e+00</th> </td></td>	POLN_HEVPA <td>NON-STRUCTURAL POLYPRO <th>3.17e+00</th> </td>	NON-STRUCTURAL POLYPRO <th>3.17e+00</th>	3.17e+00
45	91	8.2	440	1 <td>YW19_MYCTU <td>HYPOHETICAL 45.3 KD T <th>4.15e+00</th> </td></td>	YW19_MYCTU <td>HYPOHETICAL 45.3 KD T <th>4.15e+00</th> </td>	HYPOHETICAL 45.3 KD T <th>4.15e+00</th>	4.15e+00

ALIGNMENTS

RESULT	ID	IEX1_HUMAN	STANDARD;	PRT;	156 AA.
AC	P46695; Q93044;				
DT	01-NOV-1995 (REL. 32, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY				
DE	PROTEIN GLY96).				
GN	IEX1.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; PRIMATES.				
NC	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-PLACENTA;				
RX	MEDLINE: 96181295.				
RA	KONDRATYEV A.D.; CHUNG K.N.; JUNG M.O.;				
RL	CANCER RES. 56:1498-1502(1996).				
CC	[2]				
RP	PRELIMINARY SEQUENCE OF 1-106 FROM N.A.				
RC	TISSUE-PLACENTA;				
RA	HILLIER L.; CLARK N.; DUBOUE T.; ELLISTON K.; HARKINS M.; HOLMAN M.;				
RA	HUTTMAN M.; KUCABA T.; LE M.; LENNON G.; MARRA M.; PARSONS J.;				
RA	RITKIN L.; ROHLING T.; TAN F.; TREVASKIS E.; WATERSTON R.;				
RL	WILLIAMSON A.; WOHLMANN P.; WILSON R.;				
CC	SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).				
CC	-1- INDUCTION: BY RADIATION.				
CC	-1- SIMILARITY: STRONG. TO MOUSE ORTHOLOG.				
CC	-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO				
CC	FRAMESHIFTS.				
DR	EMBL; S81914; G1488385; -.				
DR	EMBL; T49531; -; NOT_ANNOTATED_CDS.				
KW	GLYCOPROTEIN; TRANSMEMBRANE.				
FT	DOMAIN 1 82				
FT	TRANSEM 99				
FT	DOMAIN 100 156				
FT	CARBOHYD 133 133				
SO	SEQUENCE 156 AA; 16973 MW; 3BD528CD CRC32;				

Query Match 98.9%; Score 1102; DB 1; Length 156;
 Best Local Similarity 98.7%; Pred. No. 5.97e-191;
 Matches 154; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB	1	MCHSRSCHPMTIIOAPPTPA	IPGRSGPPEITFDPLPEPAAAPAGRPSSGRHRK	60
OY	1	MCHSRSCHPMTIIOAPPTPA	IPGRSGPPEITFDPLPEPAAAPAGRPSSGRHRK	60

 W O R L D
 (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Msrch_un n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Thu Apr 23 05:08:39 1998; Maspar time 999.58 Seconds
 1452.649 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-799-910-9
 Description: (1-1228) from US08799910.seq
 Perfect Score: 1228
 N.A. Sequence: 1 ATGTCGACTCTCCGACGCTG.....AAAAAATCTCGAG 1228
 Comp: TACACAGTGAGGCGCTGAC.....TTTTTTTTCGAGCTC

Scoring table:
 Gap 6
 TABLE default

Nmatch STD : Dbase 0; Query 0

Searched: 354530 seqs, 591221332 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

emb153
 1:em_un 2:em_or 3:em_om 4:em_ov 5:em_pl 6:em_htg
 7:em_hum1 8:em_hum2 9:em_ba 10:em_ro 11:em_un 12:em_vl
 13:em_pat
 genbank105
 14:gb_ro 15:gb_om 16:gb_ov 17:gb_in 18:gb_pl 19:gb_ba
 20:gb_st 21:gb_vl 22:gb_ph 23:gb_sy 24:gb_un 25:gb_pat
 26:gb_htg

Statistics: Mean 11.417; Variance 7.667; scale 1.489

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	118	9.6	1938 14	MMGLY96	M.musculus gly96 mRNA.	1.49e+48
2	114	9.3	1758 14	RNPRG1	R.norvegicus PRG1 gene	3.33e+46
3	79	6.4	7218 25	166494	Sequence 14 from patent	4.27e+25
4	67	5.5	7218 25	166494	Sequence 14 from patent	1.82e+19
5	41	3.3	215 25	128278	Sequence 5 from patent	5.19e+06
6	37	3.0	10772 17	AF012089	Drosophila melanogaster	4.03e-04
7	35	2.9	10772 17	AF012089	Drosophila melanogaster	3.32e-03
8	33	2.7	215 25	128278	Sequence 5 from patent	2.61e+02
9	29	2.4	800 16	GGU41467	Gallus gallus fibroblast	1.36e+00
10	29	2.4	1125 15	SSL361BP	S.scrofa mRNA for L-36	1.36e+00
11	29	2.4	1268 15	BOVIOPPP	Bovine Inorganic Pyroph	1.36e+00
12	29	2.4	1294 17	DIRHSP70	Drosophila immtlis 70	1.36e+00
13	29	2.4	1479 14	MM26SPROT	Mus musculus mRNA for	1.36e+00
14	29	2.4	1841 16	XLU69669	Xenopus laevis nuclear	1.36e+00

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
15	29	2.4	1840 17	DDU73686	Dicotylellum discoide	1.36e+00
16	29	2.4	2116 16	RCALN	Rana catesbeiana allan	1.36e+00
17	29	2.4	3042 15	SS15455	S.scrofa mRNA for inte	1.36e+00
18	29	2.4	3217 18	ZM085494	Zea mays LON1 protease	1.36e+00
19	29	2.4	3737 14	S78355	Cyl-1-cyclin D1 (mice)	1.36e+00
20	30	2.4	3789 16	XELMYBRP1	Xenopus laevis myb-rel	3.51e-01
21	28	2.3	571 17	SCU41441	Drosophila melanogaster	3.51e+00
22	28	2.3	1404 16	S76875	casein kinase 2 alpha	3.51e+00
23	28	2.3	1435 17	DMU75652	Drosophila melanogaster	3.51e+00
24	28	2.3	1908 14	MMRNASEB4	M.musculus seb4 mRNA.	3.51e+00
25	28	2.3	2031 14	MMPTHR0M	Mouse mRNA for prothro	3.51e+00
26	28	2.3	2179 25	E08204	CDNA encoding alpha 2,	3.51e+00
27	28	2.3	2301 14	RN075903	Rattus norvegicus UDP-	3.51e+00
28	28	2.3	2624 17	AF020409	Dicotylellum discoide	3.51e+00
29	28	2.3	2628 25	166342	Sequence 1 from patent	3.51e+00
30	28	2.3	5061 17	DDU25144	Dicotylellum discoide	3.51e+00
31	28	2.3	3890 17	CEIC14C11	Caenorhabditis elegans	3.51e+00
32	28	2.3	58527 18	AB007649	Arabidopsis thaliana g	3.51e+00
33	28	2.3	289893 26	CEY47810	Caenorhabditis elegans	3.51e+00
34	27	2.2	479 14	MMRORCS	Mus musculus nuclear o	8.87e+00
35	27	2.2	714 18	BVBERTY1B	B.verrucosa Bet v 1b m	8.87e+00
36	27	2.2	861 14	AF025506	Rattus norvegicus pren	8.87e+00
37	27	2.2	979 16	S7878954	CGATA-3 (chickens), liv	8.87e+00
38	27	2.2	1444 17	HELMAMIDE	H.echinata mRNA for LW	8.87e+00
39	27	2.2	2240 17	LCU56636	Lucilia cuprina alpha	8.87e+00
40	27	2.2	2498 14	MAU43333	Mesocricetus auratus 5	8.87e+00
41	27	2.2	2842 25	A26375	RIN4 coding sequence.	8.87e+00
42	27	2.2	3016 17	DDU31631	Dicotylellum discoide	8.87e+00
43	27	2.2	6258 17	PESC03080	Plasmodium falciparum	8.87e+00
44	27	2.2	8595 17	DDU14576	Dicotylellum discoide	8.87e+00
45	27	2.2	27471 17	CEIC44C3	Caenorhabditis elegans	8.87e+00

ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	118	9.6	1938 14	MMGLY96	M.musculus gly96 mRNA.	1.49e+48
LOCUS						
DEFINITION						
ACCESSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
BASE COUNT						
ORIGIN						
JOURNAL						
MEDLINE						
FEATURES						
BASE COUNT						
ORIGIN						
Query Match						
Best Local Similarity						
Matches						
DB						

QY 1 ATGTGCACTCTGCGAGCTGCGACCGGACCATGACATCCTGACGGCCCCGACCCGGCC 60

Db 772 CCCTCCACGGCCCGGAACTCCGGCGGGCTGTGTCGCGAAATTTTCACTTCGACCT 831

QY 61 CCTCCACCATCTCCGGACCCCGGGGCTCCGGTCTGAGATCTTCACTTCGACCT 120

Db 832 CTCGGGAGGGCGGTGTGTCCACGGCGCTTGAACACTTCTCGGGGACCGAANA 891

QY 121 CTCGCGAGCCCGCGACGGCCCTGCGGGGCGCCCGCCCTCTCGGGGACCGAAG 180

Db 892 CCGAGCGAAGGTGCTCTACCTCGAGTGT 923

QY 181 CCGAGCGCGAGGTTCTACCTCGAGTGT 212

RESULT 2 RNPRG1 1758 bp DNA ROD 03-SEP-1996

LOCUS DEFINITION X. norvegicus PRG1 gene.

ACCESSION X96437

NID 91515318

KEYWORDS PRG1 gene.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Ratus.

REFERENCE 1 (bases 1 to 1758)

AUTHORS Schaffer, H., Trauzold, A., Siegel, E.G., Folsch, U.R. and Schmidt, W.E.

TITLE PRG1: a novel early-response gene transcriptionally induced by pituitary adenylate cyclase activating polypeptide in a pancreatic carcinoma cell line

JOURNAL Cancer Res. 56 (11), 2641-2648 (1996)

MEDLINE 96221139

REFERENCE 2 (bases 1 to 1758)

AUTHORS Trauzold, A.

TITLE Direct Submission

JOURNAL Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of Kiel, Schittenhelmstrasse 12, Kiel, D-24105, FRG

FEATURES

source

1..1758

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/cell_line="AR4-25"

590..1279

590..1279

590..1279

gene

/gene="PRG1"

/gene="PRG1"

BASE COUNT 358 a 475 c 470 g 455 t

ORIGIN

Query Match

Best Local Similarity 76.9%; Pred. No. 3.33e-46;

Matches 163; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

LOCUS 166494 7218 bp DNA PAT 23-DEC-1997

DEFINITION Sequence 14 from patent US 5670367.

ACCESSION 166494

NID 92724471

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 7218)

AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. Gunter.

TITLE Recombinant fowlpox virus

JOURNAL Patent: US 5670367-A 14-23-SEP-1997;

FEATURES

source

1..7218

/organism="unknown"

BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others

ORIGIN

Query Match

Best Local Similarity 6.4%; Score 79; DB 25; Length 7218;

Matches 5; Conservative 220; Mismatches 146; Indels 0; Gaps 0;

Db 1061 TGCGATYYY 1120

QY 99 TGAGATCTTCACCTCGACCTCGACCTCGCGAGCCGCGACGGCGCTGCGGGCGCCAG 158

Db 1121 YY 1180

QY 159 CGCCTCGCGGGCGACCGAAGCGGAGCGGAGGTTCTACCTCGAGTGTCCGGCG 218

Db 1181 YY 1240

QY 219 CCAGCTGCGACGTCGAGAACCGAACCGCAAGAGCTCTCTTCTGCTGCTACACAT 278

Db 1241 YY 1300

QY 279 CGTCTTCGCGAGATCCGATGCTGAGAGAGGTGTGCGCGCCGCTGCTCGAGAGA 338

Db 1301 YY 1360

QY 339 CGCCCTACGCGCGATCCGTGGCGCCACCCCTGTGCTCCGAGCCCTTAA 398

Db 1361 YY 1420

QY 399 TCTGACTCGGAGCCCTCGACTACGCTGTGACTGACTTCTCGCAGAACACC 458

Db 1421 YYYYYYYYYYYY 1431

QY 459 GGCGGCTTCT 469

RESULT 4

LOCUS 166494 7218 bp DNA PAT 23-DEC-1997

DEFINITION Sequence 14 from patent US 5670367.

ACCESSION 166494

NID 92724471

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 7218)

AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. Gunter.

TITLE Recombinant fowlpox virus

JOURNAL Patent: US 5670367-A 14-23-SEP-1997;

FEATURES

source

1..7218

/organism="unknown"

BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others

ORIGIN

Query Match

Best Local Similarity 5.5%; Score 67; DB 25; Length 7218;

Matches 9; Conservative 214; Mismatches 156; Indels 0; Gaps 0;

[illegible]

LOCUS	5	128278	215 bp	DNA	PAT	30-OCT-1996
DEFINITION	Sequence 5 from patent US 5569830.					
ACCESSION	128278					
NID	91819054					
KEYWORDS	.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 215)					
TITLE	Benneft,A., Labavitch,J.M., Powell,A. and Stotz,H.					
JOURNAL	Plant Inhibitors of fungal polygalacturonases and their use to control fungal disease					
FEATURES	Patent: US 5569830-A 5 29-OCT-1996;					
source	Location/Qualifiers					
	1..215					
BASE COUNT	15 a	8 c	25 g	26 t	141 others	
ORIGIN	/organism="unknown"					
Query Match	3.3%	Score 41:	DB 25:	Length 215:		
Best Local Similarity	18.18:	Pred. No. 5.19e-06:				
Matches	36:	Conservative	79:	Mismatches 80:	Indels 4:	Gaps
Db	18	CNDRKAKDGMTTSMTTDCCNRTMGVCDTDTTFRVNDSGHKKYSSANYNGNNVGA	77			
Cp	536	CCAGTACGCGCTGTGTTCTTGTGCGTTTCGAGTCTTTTGGGAGTCCGGGA-G	478			
Db	78	THYTTTHNVNSGADSKYTDYINASGSSSSNGCTGDKNRSGADSYGSKTLMTSRN-RTKCT	136			
Cp	477	TCACAGTTAG-AAAGCGCGCGGGGTGTTCGCGGAGAAAGTGGTCCAGAGCGTAGT	419			
Db	137	ANNVDSNNMGDASVGSDDTKKHAHKSADGKVGSKNNNDPRNN-RYGTGTSNVNNGG	195			
Cp	418	CCGAGGGGCTCCGAAGTACAGTTAAAGGCTCGAGACGGGGGACACAGGGGTGGCGCCA	359			
Db	196	GKRDVSSYANNKCCGSSC 214				
Cp	358	GGGATCGCGCGCTTAGCGGC 340				
RESULT	6	AF012089	10772 bp	DNA	INV	05-AUG-1997
LOCUS						

DEFINITION	Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.
ACCESSION	AF012089
NID	g2305220
KEYWORDS	.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryote; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 4546 to 4553) Gray, Y.H., Tanaka, M.M. and Sved, J.A. P-element-induced recombination in Drosophila melanogaster: hybrid element insertion Genetics 144 (4), 1601-1610 (1996)
JOURNAL	2 (bases 1 to 10772)
MEDLINE	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
REFERENCE	Structure of the cysteine proteinase (Cp1) gene of Drosophila melanogaster and associated mutational effects Unpublished
JOURNAL	3 (bases 1 to 10772)
REFERENCE	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
AUTHORS	Direct Submission
TITLE	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia
JOURNAL	Location/Qualifiers I. 10772
FEATURES	
Source	

gene	exon	intron	CDS
gene	exon	intron	CDS
join(872. .1000,2310. .2426,6476. .6690,6751. .7707)	/gene="Cp1"	/product="cysteine protease"	join(872. .1000,2310. .2426,6476. .6690,6751. .7462)
872. .7707	/gene="Cp1"	/number=1	/codon_start=1
872. .1000	/gene="Cp1"	/gene="Cp1"	/product="cysteine proteinase-1"
872. .1000	/gene="Cp1"	/number=1	/cds_xref="PID:92305221"
1001. .2309	/gene="Cp1"	/number=2	/translation="MRTAVLLPPLALAVQAQVSFADVNEEMWTFLELRHRYODET
2310. .2426	/gene="Cp1"	join(2328. .2426,6476. .6690,6751. .7462)	EEEEPRKIRINENKHKIAKNHOFKAGKSEFLAVNKYADLHHEFQLNKNFNLYLHG
2310. .2426	/gene="Cp1"		QLAADESGKGYFTSPAHVTLPKSDVDRGTGAVTAKDQHGSCMASSSGALAEH
2310. .2426	/gene="Cp1"		HPKRSQIVLSTEDONLVDSCSTKRYGNNGGGLMNAFRYIKDNGGIDTEKSYPIEYHII
2310. .2426	/gene="Cp1"		DSCHFNKGVYGTADRGFTPIDRGDEKRAAEVATVATGVSVATDASHESQFSEGVYI
2310. .2426	/gene="Cp1"		EPQCDQONLDHGVLVVGFETDSEGYLVANVSGTWTGDKGFIKKMLRNKENGCGIA
2310. .2426	/gene="Cp1"		ASSPLPV"
2310. .2426	/gene="Cp1"		2427. .6475
2310. .2426	/gene="Cp1"		/gene="Cp1"
2310. .2426	/gene="Cp1"		4546. .4553
2310. .2426	/gene="Cp1"		/gene="Cp1"
2310. .2426	/gene="Cp1"		/note="insertion site of P(caspar)(50C)"
2310. .2426	/gene="Cp1"		/citation[1]
2310. .2426	/gene="Cp1"		6476. .6690
2310. .2426	/gene="Cp1"		/gene="Cp1"
2310. .2426	/gene="Cp1"		/number=3
2310. .2426	/gene="Cp1"		6691. .6750
2310. .2426	/gene="Cp1"		/gene="Cp1"
2310. .2426	/gene="Cp1"		6751. .7707
2310. .2426	/gene="Cp1"		/gene="Cp1"
2310. .2426	/gene="Cp1"		join(8110. .9300,9370. .>9532)
2310. .2426	/gene="Cp1"		/product="phenylalanyl tRNA synthetase"
2310. .2426	/gene="Cp1"		join(8110. .9300,9370. .>9532)
2310. .2426	/gene="Cp1"		/note="potential orf"

```

/codon_start=1
/product="phenylalanyl tRNA synthetase"
/db_xref="PID:9230522"
/translaction="MLTLLRVOGARHWIKSTRCLASSAAPKSPSPPOLEVGSGTYA
TDGNTNYPKILSYGVANKRLHLOTDPILIIQRIIVNFEYGARNRQNPFLFSYVDKNNK
PVMTVQENDLILPADHVSROKSDCYXINOQLRRAHTTAHOVELLSGIDNLLVYG
EYVRFEIDSTHYHPVFOADAVRLVTKDXLFEENRPGLELEPETSGLADPKLPPRP
SSNRKPNSPATRAVRKLMHEMKRHYLVGLTDLGPRKIRWDTTFPFPTSPNELITI
YFKONWLEVLGCGIMRHEILQRSVHOSIGAFVGLERLAWLFDLPIDRLFNWSNDS
GFLQFSEKEDLANLPKPKPISHYPCQNDLSFMWLPDIEVDAGSPNDFYDLRSVAG
DMVQISLVDKFKHPKTKGSSVCFRIYRHEMRITLQAEVNEIKHOLASAVDSFNQ
IX"
BASE COUNT      2929 a      2357 c      2282 g      3046 t      158 others
ORIGIN
Query Match      3.0%; Score 37; DB 17; Length 10772;
Best Local Similarity 18.1%; Pred. No. 4.03e-04;
Matches 19; Conservative 57; Mismatches 27; Indels 2; Gaps 2;

Db 1776 KKTYYTMMKKWYTSRPTTTSAMMMWYTWTSWYMYANVAMKMMWTRTWAR-MAMASW 1774
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 706 CGTATTATTATTCCTATGCTCCTCAATAAATTAATTAATGATTAATTAATGCTCTCTC 765
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

Db 1775 ARMKKTSAAAYASAMRKMMKMYARANKKTMMAAMKMYRAAW 1819
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 766 AGGTATGTGAGATG-TGTACGTAAATATTATTATTACTATGCAA 809
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

RESULT 7
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete
            cds, and phenylalanyl tRNA synthetase gene, partial cds.
ACCESSION AF012089
NID 92305220
KEYWORDS
SOURCE
ORGANISM
            fruit fly.
            Drosophila melanogaster
            Eukaryota; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 4546 to 4553)
            Gray,Y.H., Tanaka,M.M. and Sved,J.A.
            P-Element-induced recombination in Drosophila melanogaster: hybrid
            element insertion
            Genetics 144 (4), 1601-1610 (1996)
            2 (bases 1 to 10772)
            Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
            Structure of the cysteine proteinase (Cp1) gene of Drosophila
            melanogaster and associated mutational effects
            Unpublished
            3 (bases 1 to 10772)
            Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
            Direct Submission
            Submitted (30-JUN-1997) School of Biological Sciences, University
            of Sydney, Biology A12, Sydney University, NSW 2006, Australia
            Location/Qualifiers
            1..10772
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            join(872..1000,2310..2426,6476..6690,6751..7707)
            /gene="Cp1"
            /product="cysteine protease"
            872..7707
            /gene="Cp1"
            872..1000
            /gene="Cp1"
            /number=1
            1001..2309
            /gene="Cp1"
            2310..2426
            /gene="Cp1"
            /number=2

```

[illegible]

```

source          1. 215
BASE COUNT      15 a      8 c      25 g      26 t      141 others
ORIGIN

Query Match      2.7%: Score 33; DB 25; Length 215;
Best Local Similarity 12.2%: Pred. No. 2.61e-02;
Matches 25; Conservative 79; Mismatches 101; Indels 0; Gaps 0;

Db 11 VVSRTASCDNRKAKDQNTSSMTTDCNNTWGCYCDTDTTYRNVNDSGHHKYSANNYCG 70
OY 799 AACTTATGCAAGGCTGTGATGCTCCCTCTGTAATGCGAGCTCTTGTGATTAT 858
Db 71 NNVGAKTHYHTVNSGADSKTYVDSYASGTSNGSDGNKSGADSYGSKATMTSR 130
OY 859 TAGACTTTGTGGAGCTGCGAGACGACACCTGGAAGTGGCGCAAGTAGAGAGAA 918
Db 131 NRTGTANNAVDSRNKGADSVGSKTKKNAKNSADGKYSKNNDRNNRYGTGTSNVS 190
OY 919 ATGGGAGAGACTCGGCTGGGGAGGAGCGCTCCGGCTGGATGAGTGTGTGGGTGCG 978
Db 191 NNGCGGNKRDVSSYANNKCGSSCT 215
OY 979 TAAGTTTAGAGGTGACTGCTACTCT 1003

RESULT 9
LOCUS          GGU41467      800 bp      mRNA
DEFINITION     Gallus gallus fibroblast growth factor 8 FGF8 mRNA, partial cds.
ACCESSION      U41467
KEYWORDS       NID
SOURCE         chicken.
ORGANISM       Gallus gallus
Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
Phasianidae; Phasianinae; Gallus.
REFERENCE      1 (bases 1 to 800)
AUTHORS        Crossley, P.H., Minowada, G., MacArthur, C.A. and Martin, G.R.
TITLE          Roles for FGF8 in the induction, initiation, and maintenance of
               chick limb development
JOURNAL        Cell 84 (1), 127-136 (1996)
MEDLINE        96140646
REFERENCE      2 (bases 1 to 800)
AUTHORS        Crossley, P.H., Minowada, G., MacArthur, C. and Martin, G.R.
TITLE          Direct Submission
JOURNAL        Submitted (27-NOV-1995) Philip H. Crossley, Anatomy, UCSF,
               Parnassus, San Francisco, CA 94143-045, USA
FEATURES       source
               1..800
               /organism="Gallus gallus"
               /db_xref="taxon:9031"
               /dev_stage="embryo"
               /tissue_type="brain"
               /clone_lib="E10 embryonic chick brain cDNA library"
               <1..615
               /codon_start=1
               /product="fibroblast growth factor 8 FGF8"
               /db_xref="PID:g1134970"
               /translation="VFMLFVLCLOAVTVQSPNFTQHVREOSLVTDLQSLRLVITY
               QLSYRSGKHVQILDNKKINAMAEEDGDVAKLIVERDFGSRIRKGAATGYICMKN
               KGLIKSGNKGKDCVETIVLENNYTAIONAKYEGMYAFTFRKGPFRKSGKTRQHOR
               EVHFMKRLPKGHOTEPHRRFEFLNPFRRSRKRTNNSARVRP"
CDS
               225 a      219 c      226 g      130 t

BASE COUNT      225 a      219 c      226 g      130 t
ORIGIN

Query Match      2.4%: Score 29; DB 16; Length 800;
Best Local Similarity 74.6%: Pred. No. 1.36e+00;
Matches 44; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 742 TGTCTGTTTACAAAAAACCACCAACAAAAAAGAAAAAAGTGCAG 800
OY 1170 TGTCTGTGAACAAATAATGATTACTGTCAAAAAAAGAAAAAAGTGCAG 1228

```

```

RESULT 10
LOCUS          SSU36LBP      1125 bp      RNA
DEFINITION     S.scrofa mRNA for L-36 lactose binding lectin.
ACCESSION      X79303
KEYWORDS       1-96LBP gene; lactose-binding lectin.
SOURCE         pig.
ORGANISM       Sus scrofa
Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euteria; Artiodactyla; Suidae; Sus.
REFERENCE      1 (bases 1 to 1125)
AUTHORS        Chiu, M.L., Parry, D.A., Feldman, S.R., Klapper, D.G. and O'Keefe, E.J.
TITLE          An adherens junction protein is a member of the family of
               lactose-binding lectins
JOURNAL        J. Biol. Chem. 269 (50), 31770-31776 (1994)
MEDLINE        95081129
REFERENCE      2 (bases 1 to 1125)
AUTHORS        O'Keefe, E.J.
TITLE          Direct Submission
JOURNAL        Submitted (18-MAY-1994) E.J. O'Keefe, University of North Carolina,
               137 NCMH, Dept of Dermatology, CB7600, Chapel Hill, NC 27514, USA
FEATURES       source
               1..1125
               /organism="Sus scrofa"
               /db_xref="taxon:9823"
               /dev_stage="adult"
               /tissue_type="epithelium, oral, tongue"
               /cell_type="keratinocyte"
               /clone_lib="pig tongue cDNA"
               49..1020
               /gene="136LBP"
               49..1020
               /gene="136LBP"
               /codon_start=1
               /product="L-36 lactose binding protein"
               /db_xref="PID:g623346"
               /translation="MAFVAPAGVQPTNPTLPYKPIPGGLGVMSVYIOGVANEHMK
               RFVNFVVGSGADVAHFENPREFDGMGVNSQDGMKMGKSRKPAPE
               LVNVLDPKHVYVYVNGDPYEFGRIPVQVLTHTLOVNGDLTLOSINFGGQAPSPG
               MPNPGYEPGKHNOQPCNLPMEGAPFNPVPYKTRLOGLVARRIVYIGYPPSG
               KSLVINFGSSGVALHINRLTGLIVRNSYLVNGKGAERSSSEVPFAPGQFTDL
               SIRCGLDLRFKYYANGOHLDFFSHLSNFGVDLTLEIGDVLTSYVQI"
BASE COUNT      255 a      347 c      303 g      220 t
ORIGIN

Query Match      2.4%: Score 29; DB 15; Length 1125;
Best Local Similarity 89.2%: Pred. No. 1.36e+00;
Matches 33; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1089 AATAAATTAATAAAGTGCACAAAAAAGAAAAA 1125
OY 1185 AATAAATTAATAAAGTGCACAAAAAAGAAAAA 1221

RESULT 11
LOCUS          BOVIOPPR      1266 bp      mRNA
DEFINITION     Bovine inorganic pyrophosphatase mRNA sequence.
ACCESSION      M95283
KEYWORDS       9163228
               inorganic pyrophosphatase.
SOURCE         Bos taurus retina cDNA to mRNA.
ORGANISM       Bos taurus
Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euteria; Artiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 1266)
AUTHORS        Yang, Z. and Wensel, T.G.
TITLE          Molecular cloning and functional expression of cDNA encoding a
               mammalian inorganic pyrophosphatase
JOURNAL        J. Biol. Chem. 267, 24641-24647 (1992)
MEDLINE        93077559

```

```

FEATURES
  SOURCE
    1..1266
    /organism="Bos taurus"
    /db_xref="taxon:9913"
    /tissue-type="retina"
  BASE COUNT
    396 a 239 c 300 g 331 t
  ORIGIN

Query Match
  Best Local Similarity 87.2%; Pred. No. 1.36e+00;
  Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1227 AATATAAATTTATTTGTTTCAAAAAAAAAAAAAAAAAAAAAA 1265
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
  Oy 1184 AATATAAATTTATTTGTTTCAAAAAAAAAAAAAAAAAAAAAA 1222

RESULT 12
LOCUS DIRHSP70 1294 bp mRNA INV 25-AUG-1992
DEFINITION Dirofilaria immitis 70 kDa heat shock protein homologue (hsp 70)
ACCESSION M95648
KEYWORDS heat shock protein 70; heat shock protein homologue.
SOURCE Dirofilaria immitis adult CDNA to mRNA.
ORGANISM Dirofilaria immitis
  Secermentae: Spirotruta; Spirotruta: Spirotruta: Filarioidea;
  Eukaryotae: Spirotruta; Spirotruta: Spirotruta: Filarioidea;
  Onchocercidae: Dirofilaria.
REFERENCE
  1 (bases 1 to 1294)
  Cuilepper, J.A., Friedman, L. and Dale, B.
  Molecular cloning and characterization of a Dirofilaria immitis
  CDNA encoding an Hsp 70 homologue
  Unpublished (1992)
FEATURES
  SOURCE
    1..1294
    /organism="Dirofilaria immitis"
    /db_xref="taxon:5287"
    /dev_stage="adult"
    1..1038
    /partial
    /note="homologue"
    /codon_start=1
    /product="heat shock protein 70"
    /db_xref="PID:g156706"
    /translation="REFELCADLFRSTMDPYEKLRAKMDKQVHDIYLVGSTRIP
    KVKLLSDPFSKGLNKSINPDEAVAGVQAALISGDSKSEAQDILLDLVAPLSIG
    IETAGVWYALIKRNTIPTKTSQFTTYSNPGVLIQVYEGEAMTKNNLKGFE
    LSGIPAPRGVPOIEVTFIDANGILNVAQDSKSTGKONITITNDKGRLSKDEIB
    VQAEKRYKADDEOKRIIAKNLESAFNMKOTIEDEKIKDKISDDKKIKKDECE
    TVRILDNQTAENDEFHROKLESCVNPITTKLYQSAQMGPMGMPGAGAGGS
    TGGGPTTEEDV"
  BASE COUNT
    409 a 229 c 296 g 360 t
  ORIGIN

Query Match
  Best Local Similarity 87.2%; Pred. No. 1.36e+00;
  Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1256 AAGTAAATTTATATGCGGTCAAAAAAAAAAAAAAAAAAAAAA 1294
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
  Oy 1184 AATATAAATTTATTTGTTTCAAAAAAAAAAAAAAAAAAAAAA 1222

RESULT 13
LOCUS MM26SPROT 1479 bp RNA ROD 07-OCT-1997
DEFINITION Mus musculus mRNA for 26S proteasome non-ATPase subunit.
ACCESSION Y13071
KEYWORDS 26S proteasome; non-ATPase subunit; proteolysis; regulatory
  complex.
SOURCE house mouse.
ORGANISM Mus musculus

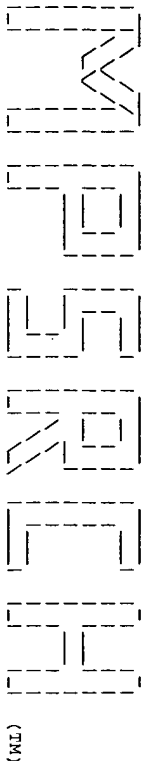
FEATURES
  SOURCE
    1..1479
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /dev_stage="adult"
    237..1166
    /gene="PADI"
    237..1166
    /gene="PADI"
    /function="proteolysis"
    /codon_start=1
    /product="26S proteasome, non-ATPase subunit"
    /db_xref="PID:g352083"
    /db_xref="PID:g3505940"
    /translation="MDRLRLGGMGPLRAPPTDAPVDAEQVYISSALLKMLKRG
    RAGVPMVGMGLGFEVDYTVIVDFVPMPOSGTGVSEAVDPVQAMLMKQTG
    RPPMYGMYHSRPGCMISGNDINQOSFEALSEAVAVYVDPISYGVKTYDIFR
    LIANMMVIGHEROTYSNGLHINKSIALHGLNRYSTITINRYKLEDEKMLN
    LHRKSMESGLTQDYSERHKNESVVKEMLEAKNNKAVEEDKMTPOLAIKNVKG
    QDPKRLLEHVDVMTSNIVOCIAAALDLYVRK"
  BASE COUNT
    453 a 270 c 349 g 407 t
  ORIGIN

Query Match
  Best Local Similarity 87.2%; Pred. No. 1.36e+00;
  Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1440 AATATAAATCATTTACATCCAAAAAAAAAAAAAAAAAAAAAA 1478
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
  Oy 1184 AATATAAATGATTACTGTCTCAAAAAAAAAAAAAAAAAAAAAA 1222

RESULT 14
LOCUS XLU69669 1811 bp mRNA VRT 26-FEB-1997
DEFINITION Xenopus laevis nuclear pore complex-associated protein TPR (tpr)
ACCESSION M96969
KEYWORDS mRNA, partial cds.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
  Eukaryotae: mitochondrial eukaryotes; Metazoa: Chordata;
  Vertebrata: Amphibia; Batrachia: Anura; Mesobatrachia: Pipidoidea;
  Pipidae; Xenopodinae; Xenopus.
REFERENCE
  1 (bases 1 to 1811)
  Cordes, V.C., Reidenbach, S., Rackwitz, H.R. and Franke, W.W.
  Identification of protein p270/TPR as a constitutive component of
  the nuclear pore complex attached intranuclear filaments
  J. Cell Biol. 136 (3), 515-529 (1997)
  7177132
  MEDLINE
  REFERENCE
  2 (bases 1 to 1811)
  Cordes, V.C., Hunzicker, A. and Franke, W.W.
  Direct Submission
  JOURNAL
  Submitted (06-SEP-1996) Cell Biology/0110, German Cancer Research
  Center, INF 280, Heidelberg 69120, Germany
  Location/Qualifiers
    1..1811
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"

```

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Apr 23 05:39:41 1998: Maspar time 157.52 Seconds

Tabular output not generated. 899.629 Million cell updates/sec

Title: >US-08-799-910-9

Description: (1-1228) from US08799910.seq

Perfect Score: 1228

N.A. Sequence: 1 ATGTCACACTCTCCGACGCTG.....AAAAAAAAAACTCGAG 1228

Comp: TACACAGTGAAGAGCGTCGAC.....TTTTTTTTTTTGAGCTC

Scoring table: TABLE default

Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 159651 seqs, 57698962 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

n-geneseg30
1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33

Statistics: Mean 9.360; Variance 6.778; scale 1.381

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	135	11.0	297	10	077534 Human genome fragment	1.11e-57
2	94	7.7	1047	2	010572 Human Natriuretic Pep	5.93e-35
3	80	6.5	1047	2	010572 Human Natriuretic Pep	2.20e-27
4	62	5.0	267	10	077554 Human genome fragment	6.80e-18
5	47	3.8	33	18	120789 Human gene signature	2.61e-10
6	46	3.7	91	9	051746 Oligonucleotide probe	8.07e-10
7	43	3.5	91	9	051746 Oligonucleotide probe	2.31e-08
8	43	3.5	204	1	N81164 Base substituted E.co	2.10e-07
9	41	3.3	204	1	N81164 Base substituted E.co	1.59e-05
10	37	3.0	114	12	070469 Generic DNA sequence	1.32e-04
11	35	2.9	114	12	070472 Generic DNA sequence	1.32e-04
12	36	2.9	114	12	070468 Generic DNA sequence	4.59e-05
13	36	2.9	114	12	070467 Generic DNA sequence	4.59e-05
14	36	2.9	114	12	070466 Generic DNA sequence	4.59e-05
15	35	2.9	178	31	176405 Human endothelin-1 an	1.32e-04

C	16	34	2.8	100	31	T76186	Human IL4 receptor an	3.74e-04
C	17	34	2.8	114	12	070465	Generic DNA sequence	3.74e-04
C	18	34	2.8	114	12	070466	Generic DNA sequence	3.74e-04
C	19	34	2.8	114	12	070470	Generic DNA sequence	3.74e-04
C	20	34	2.8	114	12	070465	Generic DNA sequence	3.74e-04
C	21	34	2.8	128	31	T76233	Human IL6 antisense o	1.05e-03
C	22	33	2.7	114	12	070469	Generic DNA sequence	1.05e-03
C	23	33	2.7	178	31	T76405	Human endothelin-1 an	2.93e-03
C	24	32	2.6	114	12	070467	Generic DNA sequence	2.93e-03
C	25	32	2.6	114	12	070468	Generic DNA sequence	2.93e-03
C	26	32	2.6	172	32	T76363	Human Interleukin 8 a	8.08e-03
C	27	31	2.5	114	12	070470	Generic DNA sequence	8.08e-03
C	28	31	2.5	2420	7	Q46672	p mysa cDNA clone.	5.92e-02
C	29	29	2.4	88	31	T76170	Human IL3 receptor an	2.20e-02
C	30	30	2.4	100	31	T76186	Human IL4 receptor an	2.20e-02
C	31	30	2.4	114	12	070473	Generic DNA sequence	2.20e-02
C	32	30	2.4	114	12	070471	Generic DNA sequence	2.20e-02
C	33	30	2.4	114	12	070472	Generic DNA sequence	2.20e-02
C	34	29	2.4	190	31	T76452	Chymase antisense o1	5.92e-02
C	35	29	2.4	250	31	T76438	Substance P antisense	1.57e-01
C	36	28	2.3	114	12	070471	Generic DNA sequence	1.57e-01
C	37	28	2.3	130	31	T76152	Human vascular cell a	1.57e-01
C	38	28	2.3	162	31	T76307	Human RANTR5 antisens	1.57e-01
C	39	28	2.3	200	31	T76398	Human leukotriene C4	1.57e-01
C	40	28	2.3	595	19	T16988	E-Dex integrin inhibi	1.57e-01
C	41	28	2.3	2179	13	Q73117	Alpha 2, 3-sialyl tra	1.57e-01
C	42	28	2.3	2232	12	Q77939	Human alpha-2,3-sialy	1.57e-01
C	43	28	2.3	2992	13	Q79934	Human liver hap cDNA	1.57e-01
C	44	28	2.3	2992	20	T06491	Human hepatoma cellno	1.57e-01
C	45	28	2.3	2992	11	Q65572	Human liver hap cDNA	1.57e-01

ALIGNMENTS

RESULT	1
ID	077534 standard; DNA; 297 BP.
AC	077534:
DT	23-SEP-1994 (first entry)
DE	Human genome fragment. (Preferred)
KW	Brain; Placenta; bone marrow; genetic analysis; gene mapping;
OS	detection; homology; human; adrenal tissue; ds.
PN	Homo sapiens.
PI	WO9401548-A.
PF	20-JAN-1994.
PR	13-JUL-1993; G01467
PA	13-JUL-1992; GB-014657.
PI	Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
DR	PIBson DR, Starkey M;
PT	WPI: 94-035056/04.
PS	New nucleic acid fragment encoding gene products - can be used
PS	for genetic analysis and mapping
PS	Claim 1: Page 575-576; 61pp; English.
CC	Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC	the placenta or bone marrow comprise any of: (A) a sequence
CC	selected from (076401-077613), (B) an allelic variation of a
CC	sequence as described in (A), or (C) a sequence complementary
CC	to (A) or (B).
CC	Preferred sequences exhibit no more than 90105mology to a human
CC	CC sequence known per se.
CC	Sequence 297 BP: 66 A; 117 C; 61 G; 52 T;
SO	Sequence
Query Match	11.0%; Score 135; DB 10; Length 297;
Best Local Similarity	76.6%; Pred. No. 1.11e-57;
Matches	209; Conservative 0; Mismatches 62; Indels 2; Gaps 2;
Db	22 cacttcgagcgtgcacaccccaatgacacccatcagaccccaacccacccacac 81
OY	7 CACTTCGACAGTCCGACCCGACCATCCTCAGAGCCCGACCCGCGCCCTCC 66
Db	82 gttatcgtccagccacaataag-citgcgcgcctagatctccacccacacccctccg 140
OY	67 ACCATCCCGGAGCCCGCGGCGCTCCGAGATCTTCACCTTCGACCCCTCCCG 126


```

TT      /*tag= a      /*function= multiple cloning site
FT      primer_bind  187..204
FT
FT      /*tag= b
PN      EP-285123-A.
PD      05-MAY-1988.
PF      30-MAR-1988.
PR      03-APR-1987; US-034819.
PT      (SUSO) SUOMEN SOKERI OY.
PI      Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
DR      WPI; 88-279927/40.
PT      Introducing random point mutations into nucleic acids -
PT      by prepn of single stranded template, annealing a primer, elongation,
PT      misincorporation, completion of molecules and screening.
PS      disclosure; P: English.
CC      Random point mutations were introduced into the alpha fragment of
CC      E. coli beta-galactosidase. The wild type sequence was obtained as a
CC      single stranded template and an oligonucleotide was hybridised to
CC      it to generate a popn of DNA molecules which terminate at all
CC      possible nucleotide positions within a specified region. The
CC      variable 3' ends generated in this way are used as primers for
CC      reverse transcriptase. Nucleotides are misincorporated by the
CC      transcriptase and the molecules are completed to forms that can be
CC      amplified and then expressed in a suitable host-vector system.
CC      The sequence covers all 176 diff base substitutions, most of which
CC      occurred singularly in any given mutant.
CC      See also P80575.
SQ      Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.5%; Score 43; DB 1; Length 204;
Best Local Similarity 14.5%; Pred. No. 2,31e-08;
Matches 18; Conservative 61; Mismatches 44; Indels 1; Gaps 1.

Db      62 ywcgagcycaayvcdehvgccgymrttthyrmydhyrdsdaawycgyrrsv 121
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CP      1134 CACCAACCCCAACCCCTTACTCTGCACACCACCCCAAGGCGACTTCAGAGA- 1076
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db      122 kydcynacchdhvvybbbyvynvnhnncocbnhvcnvhbnhnrwayrhdrr 181
      : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
CP      1075 GGATCTCAGCGATCTCATCTCTTAATGATGCGCCGAGTCTCACAGACAGCGAG 1016
      : : : : : || : : : : : || : : : : : || : : : : : || : : : : :

Db      182 ddtvh 185
      : : : : :
CP      1015 TTGA 1012

RESULT 9
ID      N81164 standard; DNA; 204 BP.
AC      N81164.
DT      08-NOV-1990 (first entry)
DE      Base substituted E. coli beta-galactosidase alpha-fragment.
KW      E. coli beta galactosidase alpha-fragment; base substitutions; ss.
OS      Escherichia coli.
FH      Key
FH      Location/Qualifiers
FT      misc-feature 19..69
FT      /*tag= a
FT      /function= multiple cloning site
FT      primer_bind 187..204
FT      /*tag= b
PN      EP-285123-A.
PD      05-MAY-1988.
PF      30-MAR-1988; 105163.
PR      03-APR-1987; US-034819.
PT      (SUSO) SUOMEN SOKERI OY.
PI      Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
DR      WPI; 88-279927/40.
PT      Introducing random point mutations into nucleic acids -
PT      by prepn of single stranded template, annealing a primer, elongation,
PT      misincorporation, completion of molecules and screening.
PS      disclosure; P: English.
CC      Random point mutations were introduced into the alpha fragment of
CC      E. coli beta-galactosidase. The wild type sequence was obtained as a
CC      single stranded template and an oligonucleotide was hybridised to

```

[illegible]

[illegible]

TSAR: totally synthetic affinity reagent; **synthetic:** binding domain;
KW effector domain; concatenated heterofunctional protein; **linker:**
KM direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.

FH Key Location/Qualifiers
FT misc.feature 55..60
FT /*tag= a
FT /note "this sequence represents 'Z'; Z can be a
sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PD MO94I8318-A.
PN 18-AUG-1994.
PF 01-FEB-1994: UO0977.
PR 01-FEB-1993: US-013416.
PR 30-DEC-1993: US-176500.
PR 31-JAN-1994: US-189331.
PA (UNC-) UNIV NORTH CAROLINA.
PI Powles DM, Kay BK:
PWI. 94-279739/34.
DR P-PDSB: R65152.
PT Identifying proteins or peptide(s) which bind a ligand - by
screening a recombinant vector library expressing fusion proteins
comprising a binding domain and an effector domain
PS Disclosure: Page 35; 255pp; English.
CC G70466 is a generic DNA sequence used to generate random TSAR (Totally
Synthetic Affinity Reagents) peptides.This generic formula can also be
represented as follows:(NNB)1(TGC)(NNB)10(TGC)2(NNB)4(NTG)(NNB)
-9Y. X and Y are flanking restriction sites (X is not the same as Y)
that are not specified further. Other generic sequences are shown in
C70466-66. Other specific peptides generated by these generic sequences
are shown in RS5131-54. TSARs are concatenated heterofunctional proteins
or peptides comprising at least two functional regions - a binding
domain with affinity for a ligand and a second effector peptide portion
that is chemically or biologically active.They may further comprise a
linker peptide between the 2 domains.The oligonucleotides are also
designed so that the expressed peptide contains 2 or 4 cysteine residues
positioned in, or flanking, the unpredicted or variant residues. These
residues confer some degree of conformational rigidity to the peptides.
CC The TSARs or compns. comprising a TSAR binding domain can be used in
vivo to deliver a chemically or biologically active moiety, eg. metal
ion, radiotope, peptide, toxin or enzyme, to the specific target or
on the cell. They can also replace the function of macromolecules, eg.
monoclonal or polyclonal antibodies and therefore circumvent the need for
complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
direct and rapid detection in a screening process.
SQ Sequence 114 BP: 0 A; 4 C; 4 G; 4 T;

Query Match 2.98; Score 36; DB 12; Length 114;
Best Local Similarity 7.38; Pred. No. 4.59e+05;
Matches 8; Conservative 31; Mismatches 71; Indels 0; Gaps 0;

Db 5 gcnnbnbnbnbnbnbnbnbnbnbnbnbnbtgcgnbnbnbnbnbnbnbnbnbnbn 64
Cc | |||
CP 703 GCCTCTTCACAGCGGCCCTAACCCTGGGTGTCCTCTATGGCCCTCG 644
Dc | |||
Db 65 nbnnbnbnbnbnbnbnbnbtgcnbnbnbnbnbnbnbnbnbnbnbnbnbn 114
Cc | |||
CP 643 GTCTCTTGCGCCGTGGTCCCCTCAAGCACCGGATGGCTCTCCGCTG 594

RESULT 15
ID T76405 standard; DNA: 178 BP.
AC T76405;
DT 15-SEP-1997 (first entry)
DE Human endothelin-1 antisense oligonucleotide.
KW Asthma; airway epithelium; adenosine free; cystic fibrosis;
KM Chronic obstructive pulmonary disease; bronchitis; ss.
OS Synethetic.
PN WO9640162-A1.
PD 19-DEC-1996.
PF 06-JUN-1996: U09305.
PR 07-JUN-1995: US-474497.

(TM)

Distribution rights by Oxford Molecular Ltd

Tabular output not generated.

Sequence:

Searched:

Statistics:

and is derived by analysis of the total score distribution.

SUMMARIES

23 98

45

ALIGNMENTS

SQ SEQUEN

Query Mat

Matches

Db 1 N

QY 1 N

Db 61 F

QY 61 F

Db 121 A

QY 121 A

100

ID S33363

AC XXXXXX

DT 01-JAN-68

DE · This :
xy

CC A;Circ

CC A; Int
CC C; Vot

SQ SEQUE

Query Ma
Post T

OY 135 TSEPS 139

RESULT 6
ID MSETQEOASGNGEPDLPPTIRVTLKTLDDREATVIGLQDTIQSLIDGRREMIQSGFORYIAGRVL STANDARD

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
SO SEQUENCE 1729 AA; 194986 MW; 14801419 CN;

Query Match 9.8%; Score 109; DB 2; Length 1729;
Best Local Similarity 23.7%; Pred. No. 8.78e+01;
Matches 31; Conservative 30; Mismatches 65; Indels 5; Gaps 5;

Db 774 FLENGRIPSTSSAPSTSENPGSPFNS-EDADIRAGRLPLGTRPNRRTVRETVHAAA 832
OY 13 ILQAPTPAPSTIPGPRRSGPEITFTDPLPEPAAPAGR-PSASGHRKRSRVLYPRVY 71
Db 833 ARAESPNIHISTFTATHTFAPAGF-PLMASSNVPSTSGPPGPIRQVVSPTPTBGL 891
OY 72 RR-QLPVEEPNPAKRLFLTLTLTVFCQILMAEGVAPALP-PEADAPNAAALAPTVSPVL 129
Db 892 FEFDLGSSDQ 902
OY 130 EFPNLTSEPSD 140

RESULT 7
ID JQ0532 STANDARD; PRT; 753 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
XX
DE
XX
CC A:Accession: JQ0532
CC A:Molecule type: mRNA
CC A:Residues: 1-753 <DIN>
CC A:Experimental source: strain Jervis Bay isolate
SO SEQUENCE 753 AA; 82427 MW; 3187938 CN;

Query Match 9.2%; Score 102; DB 2; Length 753;
Best Local Similarity 32.9%; Pred. No. 3.67e+00;
Matches 24; Conservative 14; Mismatches 30; Indels 5; Gaps 5;

Db 271 LPRPGNPGVLPQPKYRGA-PSS-NL-PLPTQAPPRARELQRSILHLASRONAPRLRP 327
OY 14 LQAPTPAPSTIPGPR-RGSGPEITFTDPLPEPAAPAGRPSASGHRKRSRVLYPRVY 72
Db 328 RRLR-SDPIQOTR 339
OY 73 ROLPVEEPNPAK 85

RESULT 8
ID S20590 STANDARD; PRT; 913 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
CC TOIG of: s20590 check: 4303 from: 1 to: 913
CC >P1:S20590
CC exo-alpha-stalidase (EC 3.2.1.18) - Actinomyces viscosus

CC C:Species: Actinomyces viscosus
CC C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 09-Sep-1997
CC C:Accession: S20590
CC R:Henningsen, M.; Roggentin, P.; Schauer, R.
CC Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991
CC C:A:Title: Cloning, sequencing and expression of the stalidase gene from Actin

CC C:A:Reference number: S20590; M01D:92162190
CC C:A:Accession: S20590
CC C:A>Status: preliminary
CC C:A:Molecule type: DNA
CC C:Residues: 1-913 <HEI>
CC C:A:Cross-references: EMBL:X62276; NID:q39254; PID:q39255
CC C:Keywords: glycosidase; hydrolase
SO SEQUENCE 913 AA; 96216 MW; 3970532 CN;

Query Match 9.2%; Score 103; DB 2; Length 913;
Best Local Similarity 29.9%; Pred. No. 3.00e+00;
Matches 20; Conservative 14; Mismatches 31; Indels 2; Gaps 2;

Db 816 SPASRNAAPTPKGMPEDEID-RPSDGTMAOPTGAP-ARRVPRRRRRRPAAGCLARQ 873
OY 16 APTAPSTIPGPRRSGPEITFTDPLPEPAAPAGRPSASGHRKRSRVLYPRVYRQL 75
Db 874 RAADPGP 880
OY 76 PVEEPNP 82

RESULT 9
ID JQ0405 STANDARD; PRT; 1106 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
XX
DE
XX
CC A:Accession: JQ0405
CC A:Molecule type: DNA
CC A:Residues: 1-1106 <SHI>
CC A:Cross-references: EMBL:X15867
CC A:Note: all the codons between two in-frame stop codons are translated; the
CC A:Note: the gene encoding this protein overlaps uvra gene
SO SEQUENCE 1106 AA; 119484 MW; 4705861 CN;

Query Match 9.2%; Score 102; DB 2; Length 1106;
Best Local Similarity 29.9%; Pred. No. 3.67e+00;
Matches 20; Conservative 16; Mismatches 29; Indels 2; Gaps 2;

Db 569 PGPAVPGRRPRVPOPAAGRHPLRRRGA-HPGCHDRLRAGRPPLRRRAVHPAPA 627
OY 19 PAPSTIPGPRRSGPEITFTDPLPEPAAPAGRPSASGHRKRSRVLYPR-VVRQLPV 77
Db 628 GQPEPRH 634
OY 78 EEPNPAK 84

RESULT 10
ID S24407 STANDARD; PRT; 1206 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
CC TOIG of: s24407 check: 501 from: 1 to: 1206
CC >P1:S24407
CC formin isoform IV - mouse
CC C:Species: Mus musculus (house mouse)
CC C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997

MPEP
(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Apr 17 13:39:01 1998; Maspar time 12.21 Seconds
538.123 Million cell updates/sec
Tabular output not generated.

Title: >US-08-799-910-10
Description: (1-156) from US08799910.pep
Perfect Score: 1114
Sequence: 1 MCHSRCHPTMTILQAPTPA.....EPDVALDLSTFLQHPAAAF 156

Scoring table: PAM 150
Gap 11

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

spiremb15
1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_organelle 7:sp_phage 8:sp_plant
9:sp_bacteria 10:sp_fodent 11:sp_virus 12:sp_vertebrate
13:sp_unclassified

Statistics: Mean 42.617; Variance 98.163; scale 0.434

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1114	100.0	156	2	092691	8.46e-174
2	1114	100.2	351	11	085302	2.07e-02
3	112	10.1	115	9	P95306	3.65e-02
4	112	10.1	285	1	P78977	4.84e-02
5	111	10.0	3247	11	065553	6.41e-02
6	110	9.9	228	11	084630	6.41e-02
7	110	9.9	228	8	043558	1.12e-01
8	108	9.7	552	8	096343	5.74e-01
9	102	9.2	418	3	001662	5.74e-01
10	102	9.2	568	3	027212	5.74e-01
11	102	9.2	640	11	084171	5.74e-01
12	102	9.2	680	11	098187	4.39e-01
13	103	9.2	913	9	059164	4.39e-01
14	103	9.2	1711	9	P96311	7.51e-01
15	101	9.1	1426	10	P70298	9.79e-01
16	100	9.0	185	8	022194	9.79e-01
17	100	9.0	333	9	052486	9.79e-01
18	100	9.0	439	8	042421	9.79e-01
19	100	9.0	1048	10	063627	9.79e-01
20	100	9.0	1491	12	091718	9.79e-01

RESULT ID	1	PRELIMINARY	PRT	156 AA.	
AC	092691				BETA-GALACTOSIDASE (FRAGMENT)...
DT	01-FEB-1997 (TREMBLREL. 02, CREATED)				WP6 PRECURSOR.
DT	01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)				ALPHA-1 TYPE II COLLAG
DT	01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)				BACTINECIN 11.
DE	PGI PROTEIN (DIF-2 PROTEIN).				GENOME, PARTIAL SEQUEN
GN	PGI OR DIF-2.				PROTEIN-TYROSINE PHOSP
OC	HOMO SAPIENS (HUMAN).				ALR.
OC	EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:				ALR.
RN	ETHIOPIA: PRIMATES.				HYPOTHETICAL 40.2 KD P
RN	SEQUENCE FROM N.A.				FORK HEAD PROTEIN.
RA	TRAUZOLD A.;				BMP1.
RL	SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.				WIP1.
RP	SEQUENCE FROM N.A.				N-METHYL-D-ASPARTATE R
RA	PIETSCH A., BUECHER C., ASLANDIS C., SCHMITZ G.;				ZEIN.
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 235:4-9(1997).				CAPSID SCAFFOLD PROTEI
RP	SEQUENCE FROM N.A.				MLO PROTEIN.
RA	CANCER RES. 56:1498-1502(1996).				3-PHOSPHOINOSITIDE DEP
DR	EMBL; Y14551; E333102; -				CHITINASE PRECURSOR.
DR	EMBL; X96438; E350480; -				GABA/NORADRENALINE TRA
DR	SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;				OLEOSIN.
DR	SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;				OLEOSIN.
DR	SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;				(VSP-3) PRECURSOR.
DR	SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;				MOB3 ORF.
DR	SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;				F47B8.5.

Query Match 100.0%; Score 1114; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 8.46e-174;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB	1	MCHSRCHPTMTILQAPTPASTIPGRSGPEITFDPLPEPAAPACRPSASRGHRK	60
DB	1	MCHSRCHPTMTILQAPTPASTIPGRSGPEITFDPLPEPAAPACRPSASRGHRK	60
DB	61	RSRVLVPRVVRRLPVEEPARARLLFLITTYFCOILAAEEGVAPLPEDAPNASTL	120
DB	61	RSRVLVPRVVRRLPVEEPARARLLFLITTYFCOILAAEEGVAPLPEDAPNASTL	120
DB	61	RSRVLVPRVVRRLPVEEPARARLLFLITTYFCOILAAEEGVAPLPEDAPNASTL	120
DB	121	APTVPSPVLEPNTLSEPSDVALDLSTFLQHPAAAF	156
DB	121	APTVPSPVLEPNTLSEPSDVALDLSTFLQHPAAAF	156
DB	121	APTVPSPVLEPNTLSEPSDVALDLSTFLQHPAAAF	156

```

RESULT 2
ID 085302 PRELIMINARY; PRT; 351 AA.
AC 085302;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE HOMOLOGUE OF RETROVIRAL PSEUDOPROTEASE.
OC ORF VIRUS.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;
OC PARAPOXVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90266454.
RA FRASER M., HILL D.F., MERCER A.A., ROBINSON A.J.;
RL VIROLOGY 176:379-389(1990).
DR EMBL: M30023; G332564; -.
SQ SEQUENCE 351 AA; 39886 MW; 8EEF741B CRC32;

Query Match
Best Local Similarity 10.2%; Score 114; DB 11; Length 351;
Matches 24; Conservative 12; Mismatches 35; Indels 2; Gaps 2;

Db 236 QPRRRPRAAGARCGAPROQOPVQRAAAQOROROPRRVARRARAR 295
QY 15 QAPTPAPSTIPGRRSGGEIFTFDLPPE-PAAAPAGRPASGRHKSRRVLYPRV-VR 72
Db 296 ROORAHORRRGR 308
QY 73 ROLPVEEPNPAKR 85

RESULT 3
ID P95306 PRELIMINARY; PRT; 115 AA.
AC P95306;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE PREDICTED ORF.
OS MYCOBACTERIUM TUBERCULOSIS.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA TYAGI J.S.;
RL SUBMITTED (DEC-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA VERMA A., DASGUPTA N., AGGARWAL A.N., PANDE J.N., TYAGI J.S.;
RL INDIAN J. BIOCHEM. BIOPHYS. 32:429-436(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA TYAGI J.S.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: X63508; E282228; -.
SQ SEQUENCE 115 AA; 12106 MW; AF306566 CRC32;

Query Match
Best Local Similarity 10.1%; Score 112; DB 9; Length 115;
Matches 16; Conservative 23; Mismatches 22; Indels 2; Gaps 2;

Db 8 PSQALPPRPPTAPPPAPPPAPPLDSSRLPTAPSPRRKPNPPAPARRRRTAALR 67
QY 9 PTMTILOAPSTIPGRRSGGEIFTFDPP-LPEPAAAGRPDA-SRGHRRSRRLV 66
Db 68 YRR 70
QY 67 YPR 69

RESULT 4
ID P78977 PRELIMINARY; PRT; 285 AA.

```

```

AC P78977;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELL WALL PROTEIN PRECURSOR.
OS YAROMWIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CX39-74A; TISSUE-PINA240;
RX MEDLINE: 97127825.
RA RAMON A., GIL R., BURGAL M., SENTANDREU R., VALENTIN E.;
RL YEAST 12:1535-1548(1996).
DR EMBL: 281006; E274837; -.
KW SIGNAL.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 285 AA; 30036 MW; 9AFCAB87 CRC32;

Query Match
Best Local Similarity 10.1%; Score 112; DB 1; Length 285;
Matches 43; Conservative 29; Mismatches 70; Indels 9; Gaps 9;

Db 122 ICHTIVTLSTEVIITPPVPDQTTTAVPRPTEVEVKEPEPTPEV-PGVKPEPTRCP 180
QY 1 MCHSRSCPTMTILOAPTPAD-STIPGRRSGGEIFTFDLPPEAAPAGRPASGRHR 59
Db 181 APKPEPEVPEVKEPEPTPEVPEVREPTAPRPRL-RSLRSSPSLPLPLPPSPD-S 238
QY 60 KRSRRVLYPRVYRQLP-VEEPPNPAKRLLFLTLTVFCQIMAEQVPA-PLPEDAPNA 117
Db 239 LSLPPRSPSSSLPLSLPTLPTSL-LS 268
QY 118 ASLAP-TP-VSPVLEPFLITSEPSDYALDS 146

RESULT 5
ID Q65553 PRELIMINARY; PRT; 3247 AA.
AC Q65553;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE UL36.
GN UL36.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-COOPER;
RA SCHWYZER M., VLCEK C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.,
RA THIRY E., PACES V.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-179 FROM N.A.
RX STRAIN-COOPER;
RA SCHWYZER M., STYGER D., VOGT B., LOWERY D.E., SIMARD C.,
RA LABOISSIERE S., MISRA V., VLCEK C., PACES V.;
RL VET. MICROBIOL. 0:0-0(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-COOPER;
RA SCHWYZER M.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RX STRAIN-JURA;
RA SCHWYZER M., STYGER D., VOGT B., LOWERY D.E., SIMARD C.,
RA LABOISSIERE S., MISRA V., VLCEK C., PACES V.;
RL VET. MICROBIOL. 53:67-77(1996).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN-JURA;
RA SCHWYZER M., VLCEK C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.;

```

```

FT CHAIN 8 228 PROLINE RICH PROTEIN.
SQ SEQUENCE 228 AA: 22480 MW: 2A2B03B9 CRC32:

Query Match 9.9%: Score 110: DB 8: Length 228:
Best Local Similarity 24.0%: Pred. No. 6,41e-02:
Matches 30: Conservative 36: Mismatches 56: Indels 3: Gaps 3:

Db 36 PPPPAPPTPTTQASPPVQVOST-SPPVQSSPPVQSSPPAPPTPTTQVSSPPVQSSAPP 94
OY 16 APTAPSTTIGPRKGSSEPIFTFDLPPEPAAAGRPSARGKRSRRLVRYRVRRDL 75
Db 95 PVQSSPPPTLTTPPVQVOST-PPPSAPPPASPPFSSPPAPPTPTTQVSSPPVQSSAPP 153
OY 76 PVPEPNAKRLFLTLTIVFCQLIMAEQVAPL-PEDAPNAAASLAPTVSVLEPFLN 134
Db 154 SSPPA 158
OY 135 TSEPS 139

RESULT 8
ID Q96343 PRELIMINARY: PRT; 552 AA.
AC Q96343;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DE 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE MYOSINASE-BINDING PROTEIN RELATED PROTEIN (FRAGMENT).
OS BRASSICA NAPUS (RAPE).
OC EUKARYOTA: PLANTA: EMBRYOPHYTA: ANGIOSPERMAE: DICOTYLEDONEAE:
OC CAPPARALES: CRUCIFERAE.
RN [1].
RP SEQUENCE FROM N. A.
RC STRAIN-20516 OF SVALOF'S KARAT.
RA TATPALENSUD J., FALK A., EK B., RASK L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U59446; GI655830; -.
FT NON TER 1
SQ SEQUENCE 552 AA: 59270 MW: 6861EF25 CRC32:

Query Match 9.7%: Score 108: DB 8: Length 552:
Best Local Similarity 43.5%: Pred. No. 1.12e-01:
Matches 20: Conservative 11: Mismatches 12: Indels 3: Gaps 3:

Db 235 LTPPAPASPPAGPAPAPAGSGSH-APAPAPAPAGGGRAPAPAG 339
OY 14 LQAPTPAPSTIPGRKSGSEIFTFDLPPEPAAAP-A-G-RPSASRG 57

RESULT 9
ID Q01662 PRELIMINARY: PRT; 418 AA.
AC Q01662;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
DE SIMILAR TO CUTICULAR COLLAGEN.
DE T8F2.6.
GN CAENORHABDITIS ELEGANS.
OS EUKARYOTA: METAZOA: ACOELOMATES: NEMATODA: SECERNENTEA: RHABDITIDA.
OC [1].
RN RP SEQUENCE FROM N. A.
RC STRAIN-BRISTOL N2;
RA MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J.,
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,
RA FULTON L., GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M.,
RA JOHNSON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA LATEILLE P., LITCHING J., LLOYD C., MCMURRAY A., MORTIMORE B.,
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIEKEN L., ROOPRA A.,
RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONHAMMER E.,
RA STADEN R., SUJOSTON J., THIRRY-MIEG J., THOMAS K., VAUDIN M.,
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLIDMAN P.,

```

```

RL NATURE 368:32-38(1994).
RN (12)
RC SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RA MADSEN C., FRONICK B.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (13)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF000198; G2047346;
SO SEQUENCE 418 AA; 38279 MW; 2587D236 CRC32;

Query Match
Best Local Similarity 37.9%; Score 102; DB 3; Length 418;
Matches 22; Conservative 17; Mismatches 15; Indels 4; Gaps 3;

Db 363 AAAGAAAPPAPAAAAPAPAP-APAPAPPAPAGSGPTGVRKKVRVL--RIVR 417
QY 16 APTPAPSTIGPRGSGPELFTDPEPAAAPAGPSASRG-HRKRSRVLYPRVVR 72

RESULT 10
ID 027212 PRELIMINARY; PRT: 568 AA.
AC 027212;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE ARTICULIN P60.
OS PSEUDOMICROTHORAX DUBIUS.
OC EUKARYOTA; MITOCHONDRIAL EUKARYOTES; ALVEOLATA; CILIOPHORA; CILIATA;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-NSB;
RX MEDLINE: 96032269.
RA HUTTENLAUCH I., GEISLER N., PLESSMANN U., PECK R.K., WEBER K.,
RA STICK R.;
RL J. CELL BIOL. 130:1401-1412(1995).
DR EMBL: L41557; G767685;
SQ SEQUENCE 568 AA; 61110 MW; CA5BB376 CRC32;

Query Match
Best Local Similarity 25.0%; Score 102; DB 3; Length 568;
Matches 30; Conservative 24; Mismatches 62; Indels 4; Gaps 4;

Db 332 VNVVPVPIEVPVDRDVPVPEQLN-IDVPVDVPAARVPVERITIQGRIPLQPRVLVQ 390
QY 14 LQAPTPAPSTIGPRGSGPELFTDPEPAAAPAGPSASRGHRKRSRVLYPRVVR 73

Db 391 HVPHPVPAOEVIVQOPFAVPOPYTVQOEVPIHPVPEVPOPYAVP-QPVPVPPVAVP 449
QY 74 QLPVEEPNPAKRLLELLTIVFCQILMAEGVPAPLP-PEDAPNMASTLAPTV-SVLEVP 131

RESULT 11
ID 084171 PRELIMINARY; PRT: 640 AA.
AC 084171;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE VACCINIA VIRUS GENE F12L HOMOLOG.
OS ORF VIRUS.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-NZ2;
RA MERCER A.A.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (2)
SQ SEQUENCE FROM N.A.

RA SULLIVAN J.T., FLEMING S.B., ROBINSON A.J., MERCER A.A.;
RL VIRUS GENES 11:21-29(1995).
DR EMBL: U34774; G1002997;
DR EMBL: S82833; G1754760;
SO SEQUENCE 640 AA; 70097 MW; 0E08D7DD CRC32;

Query Match
Best Local Similarity 35.7%; Score 102; DB 11; Length 640;
Matches 20; Conservative 9; Mismatches 24; Indels 3; Gaps 3;

Db 436 RCRLVPLPLREVASQLYPGKPRMWSKRLGLIALLCVTSAAEYKIPFLDPAD 491
QY 61 RSRVLYLP-RVVARQL-PVEEPPNPAKRLLELLTIVFCQILMAEG-VPAPLPED 113

RESULT 12
ID 098187 PRELIMINARY; PRT: 680 AA.
AC 098187;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE MC018L.
GN MC018L.
OS MOLUSCUM CONTAGIOSUM VIRUS SUBTYPE 1 (MCV1).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;
RN (1)
RP SEQUENCE FROM N.A.
RA SENKEVICH T.G., BUGERT J.J., SISLER J.R., KOONIN E.V., DARAI G.,
RA MOSS B.;
RL SCIENCE 273:813-816(1996).
RN (2)
RP SEQUENCE FROM N.A.
RA SENKEVICH T.G., BUGERT J.J., SISLER J.R., KOONIN E.V., DARAI G.,
RA MOSS B.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U60315; G1491961;
SQ SEQUENCE 680 AA; 72398 MW; 43AB9271 CRC32;

Query Match
Best Local Similarity 36.0%; Score 102; DB 11; Length 680;
Matches 27; Conservative 19; Mismatches 23; Indels 6; Gaps 5;

Db 217 ACAPASAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPARTKRRGTPPLPR 276
QY 6 SHPMTLLOAPTPAPSTIGPRGSGPELFTDPEPAAAPAG-RESASGRHRR 61

Db 277 SKRV-ARAVR-QLP 289
QY 62 SRRVLYPRVVRQLP 76

RESULT 13
ID 059164 PRELIMINARY; PRT: 913 AA.
AC 059164;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE SIALIDASE (EC 3.2.1.18) (EXO-ALPHA-SIALIDASE) (NEURAMINIDASE)
DE (N-ACYLNEURAMINATE GLYCOHYDROLASE) (ALPHA-NEURAMINIDASE).
GN NANH.
OS ACTINOMYCES VISCOSUS.
OC PROKARYOTA; FIRMICUTES; IRREGULAR ASPOROGENOUS RODS; CORINEFORM GROUP.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-DSM43798;
RA HENNINGSSEN M.;
RL SUBMITTED (SEP-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-DSM43798;
RX MEDLINE: 92162190.
RA HENNINGSSEN M., ROGGENTIN P., SCHAUER E.R.;

```

B10L. CHEM. HOPPE-SEYLER 372:1065-1072(1991).
CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACTINEURAMINYL
RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED
NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
GLYCOLIPIDS OR COLOMINIC ACID.
CC EMBL: X62276; G39255;
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 913 AA; 96216 MW; 3C5A10ED CRC32;

Query Match 9.2%; Score 103; DB 9; Length 913;
Best Local Similarity 29.9%; Pred. No. 4.39e+01;
Matches 20; Conservative 14; Mismatches 31; Indels 2; Gaps 2;

Dd 816 SPASRNAAPTKPGMEPPDEID-RPSDGTAQPTGAD-ARRVPRRRRRRPAAGCLARQ 873
Oy 16 APTPASPSTPGRGSGPEIFTFDDLPPEAAAPAAGRSPASARGHRKRSRVLYPVVRQL 75

Dd 874 RAADGP 880
Oy 76 PVEEFP 82

RESULT 14 PRELIMINARY; PRT; 1711 AA.

ID P96311.
AC P96311.
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DI 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE ENOGUCCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE A)
DE (CELLULOSE A).
GN CELA.
OS AMEROCELLUM THERMOFILUM.
OC EUKARYOTIC; FIRMICUTES; LOW G+C GRAM-POSITIVE BACTERIA;
OC CLOSTRIDACEAE; CELLULOLYTIC THERMOPHILE GROUP; AMEROCELLUM.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-Z-1320.
RC ZVERLOV V.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DDJ DATA BANKS.
CC -I- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
ENOGUCCANASE ACTIVITY ON CARBOHYDRATE CELLULOSE. THE C-TERMINAL
DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CELLULOSE (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: ENDOTHERMOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
CC -I- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY E
(FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -I- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULOSE FAMILY L
(FAMILY 48 OF GLYCOSYL HYDROLASES).
CR EMBL: 286105; E350354;
DR PROSITE: PS00592; GLYCOSYL-HYDROL-F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL-HYDROL-F9_2; 1.
FT CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL; REPEAT.
FM NON_TER 1
SQ SEQUENCE 1711 AA; 189979 MW; 475389DE CRC32;

Query Match 9.2%; Score 103; DB 9; Length 1711;
Best Local Similarity 34.0%; Pred. No. 4.39e+01;
Matches 16; Conservative 15; Mismatches 15; Indels 1; Gaps 1;

Dd 630 PRTPVTPTPTPTSTAATPTPTPTTV-TPTPTPTPTAPTSTPTPT 675
Oy 9 PTWTIQAPTPTASTPTGTGSRGSCEIFTFDPLPEPAAPAGRAPSA 55

RESULT 15 PRELIMINARY; PRT; 1426 AA.

ID PT0298
AC PT0298;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DI 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CUT-LIKE 2 (CUX-2).

[illegible]

 WIREIMAGE

 (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

chmn n.a. - n.a. database search, using Smith-Waterman algorithm

on: Thu Apr 23 05:25:50 1998; Maspar time 800.19 Seconds

Tabular output not generated. 1292.427 Million cell updates/sec

Title: >US-08-799-910-9

Description: (1-1228) from US08799910.seq

Perfect Score: 1228

N.A. Sequence: 1 ATGTGTACTCTCGACGCTG.....AAAAAAAAAACTCGAG 1228

Comp: TACACAGTGAAGAGCGCGAC.....TTTTTTTTTTTGAGCTC

Scoring table: TABLE default

Gap 6

Mmatch STD: Dbase 0; Query 0

Searched: 1126798 seqs, 421087984 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

embl-est 1:em-est1 2:em-est2 3:em-est3 4:em-est4 5:em-est5

6:em-est6 7:em-est8 8:em-est9

Database:

9:gb-est1 10:gb-est2 11:gb-est3 12:gb-est4 13:gb-est5

14:gb-est6 15:gb-est7 16:gb-est8 17:gb-est9 18:gb-est10

19:gb-est11 20:gb-est12 21:gb-est13 22:gb-est14 23:gb-est15

Statistics: Mean 11.559; Variance 3.326; scale 3.476

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	495	40.3	553	12 N32077	yw97d08.s1 Homo sapien	0.00e+00
C 2	450	36.6	459	14 W60982	zc98b07.s1 Pancreatic	0.00e+00
C 3	448	36.5	470	14 AA047094	zw76c10.s1 Soares preg	0.00e+00
C 4	422	34.4	427	23 AA457705	zk59b01.s1 Soares preg	0.00e+00
C 5	420	34.2	460	19 AA043722	zk59b01.s1 Soares preg	0.00e+00
C 6	413	33.6	463	22 AA410666	zc30g09.s1 Soares ovar	0.00e+00
C 7	406	33.1	447	15 W52268	zc46a10.s1 Soares preg	0.00e+00
C 8	402	32.7	481	19 AA034911	zc46a10.s1 Soares preg	0.00e+00
C 9	383	31.2	413	15 W52369	zw91h09.s1 Homo sapien	0.00e+00
C 10	376	30.6	390	12 N29782	zw91h09.s1 Homo sapien	0.00e+00
C 11	373	30.4	443	15 W47587	zc35b02.s1 Soares sene	0.00e+00
C 12	368	30.0	475	15 W47540	zc35b02.s1 Soares sene	0.00e+00
C 13	351	28.6	355	14 AA047184	zw76c10.s1 Soares preg	0.00e+00
C 14	345	28.1	405	9 T49532	zw76c10.s1 Homo sapien	0.00e+00
C 15	332	27.0	413	18 AA301628	EST14813 Aorta endothe	0.00e+00

C 16	328	26.7	398	12 N25070	yw40f06.s1 Homo sapien	0.00e+00
C 17	327	26.6	330	13 W45220	zc98b07.s1 Pancreatic	0.00e+00
C 18	325	26.5	389	13 W39753	zc98b07.s1 Pancreatic	0.00e+00
C 19	312	25.4	398	12 N25911	yw79g12.s1 Homo sapien	0.00e+00
C 20	306	24.9	353	12 W45391	yw97d08.s1 Homo sapien	0.00e+00
C 21	268	21.8	329	9 T49531	zw76c10.s1 Soares test	0.00e+00
C 22	267	21.7	403	22 AA431547	zw76c10.s1 Soares test	0.00e+00
C 23	264	21.5	265	12 H80074	yw80g11.s1 Homo sapien	0.00e+00
C 24	255	20.8	260	19 AA372266	EST4215 Colon adenoca	0.00e+00
C 25	205	16.7	212	12 H80073	yw80g11.s1 Homo sapien	1.30e-238
C 26	189	15.4	441	15 AA080225	mj99d08.s1 Soares mous	7.67e-216
C 27	181	14.7	184	13 N57203	yw91h09.s1 Homo sapien	1.68e-204
C 28	178	14.5	463	18 AA276840	vc46a08.s1 Soares mous	2.95e-200
C 29	177	14.4	207	19 AA043424	zk57b11.s1 Soares preg	7.55e-199
C 30	177	14.4	343	14 AA003688	mg6fa09.s1 Soares mous	7.65e-199
C 31	169	13.8	182	19 AA356940	EST55572 Utricle 1-cell	1.49e-187
C 32	148	12.1	300	23 AA450645	vf80c04.s1 Soares mous	4.15e-158
C 33	136	11.1	146	19 AA034978	zk25c03.s1 Soares preg	1.97e-141
C 34	135	11.0	297	13 HSAACMB	H. sapiens putatively	4.78e-140
C 35	127	10.3	385	14 W77628	me6fa09.s1 Soares mous	5.23e-129
C 36	120	9.8	425	17 AA122977	mg23c12.s1 Barstead MP	2.10e-119
C 37	118	9.6	150	9 T10460	hdc288 Homo sapiens CD	1.14e-116
C 38	107	8.7	116	22 AA431094	zw76c10.s1 Soares test	9.98e-102
C 39	72	5.9	443	14 W71619	me40d11.s1 Soares mous	1.30e-55
C 40	65	5.3	426	17 AA073253	mm95f04.s1 Stratagene	8.88e-47
C 41	62	5.0	267	13 HSAACKHB	H. sapiens putatively	4.75e-43
C 42	57	4.6	79	12 N34155	yx79g03.s1 Homo sapien	6.34e-37
C 43	49	4.0	50	23 HMG502013	Human Hf60 3 directed	2.10e-27
C 44	43	4.0	162	22 C06218	similar to none.	2.10e-27
C 45	34	2.8	192	20 HDM4S1251	Human Chromosome 4 (cl	6.39e-11

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS
1	N32077	yw97d08.s1 Homo sapiens	CDNA clone 260175 3'	EST	10-JAN-1996
	N32077				
	91152476				
	EST.				
	human clone-260175 primer-m13 -40 forward library=Soares placenta				
	8tc9weeks 2NHP8tc9w vector-p173D (Pharmacia) with a modified				
	polylinker host-D10B (ampicillin resistant) RseI1-Not I				
	RseI2-Eco RI two placentae: one from 8 weeks and another from 9				
	weeks post conception. 1st strand cDNA was primed with a Not I -				
	15'-TGTTCACCACTGACGCGAGCGCGCGATTTTCTTTTCTTT-3') ,				
	double-stranded cDNA was size selected, ligated to Eco RI adaptors				
	(Pharmacia), digested with Not I and cloned into the Not I and Eco				
	RI sites of a modified pT73 vector (Pharmacia) library				
	constructed by Bento Soares and M. Fatima Bonaldo.				
	Homo sapiens				
	Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata;				
	Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;				
	Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;				
	Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.				
	1 (bases 1 to 553)				
	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,				
	Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,				
	Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,				
	Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and				
	Wilson,R.				
	The Washu-Merck EST Project				
	Unpublished (1995)				
	CONTACT: Wilson RK				
	Washu-Merck EST Project				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
	Tel.: 314 286 1800				
	Fax: 314 286 1810				
	Email: est@watson.wustl.edu				
	High quality sequence steps: 416				

Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Location/Qualifiers
1. .553
/organism="Homo sapiens"
/clone="260175"

BASE COUNT 153 a 165 c 106 g 127 t 2 others
ORIGIN

Query Match 40.3% Score 495; DB 12; Length 553;
Best Local Similarity 96.4%; Pred. No. 0.00e+00;
Matches 528; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

Db 5 GACAGTAATCAATTTTATTTGTTTCACAGAACATAGGCGATGATAGTGGCTC 64
|||||
Cp 1204 GACAGTAATCAATTTTATTTGTTTCACAGAACATAGGCGATGATAGTGGCTC 1145
Db 65 CGTGACAGCCCAACCCCAACCCCTTACTGCGACCCCACTAAAGGCGACTTCAAG 124
1144 CGTGACAGCCCAACCCCAACCCCTTACTGCGACCCCACTAAAGGCGACTTCAAG 1085
125 AAGATGAGAGATTCAGAGATTCATTCCTAATGATGCGGAGAGTCTCACAGAGTAGA 184
|||||
Cp 1084 AAGATGAGAGATTCAGAGATTCATTCCTAATGATGCGGAGAGTCTCACAGAGTAGA 1025
Db 185 CAGACGAGATTGAGATGCTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 244
|||||
Cp 1024 CAGACGAGATTGAGATGCTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 965
Db 245 ACTTATCCCAAGCGGAGAGTCTCCCAACCCGAGTCTCCCACTTTCTTCTCTACTT 304
Cp 964 ACTTATCCCAAGCGGAGAGTCTCCCAACCCGAGTCTCCCACTTTCTTCTCTACTT 905
Db 305 TGCCGAGATCCAGGTGCTGCTCTCCACAGTCCCAACCCGAGTCTCCCACTTTCTTCT 364
|||||
Cp 904 TGCCGAGATCCAGGTGCTGCTCTCCACAGTCCCAACCCGAGTCTCCCACTTTCTTCT 845
Db 365 ACCTGATTTTACAGCAGGAGGAGACATCTCACACCCCTGATTAATTAATTAATTAAT 424
|||||
Cp 844 ACCTGATTTTACAGCAGGAGGAGACATCTCACACCCCTGATTAATTAATTAATTAAT 785
Db 425 GTACACATCTCCATCCTAGAGAGAGTCAATTAATTAATTAATTAATTAATTAATTA 484
|||||
Cp 784 GTACACATCTCCATCCTAGAGAGAGTCAATTAATTAATTAATTAATTAATTAATTA 725
Db 485 AAG--GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 542
|||||
Cp 724 AATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 666
543 GCCTCGGT 550
|||||
Cp 665 GCCTCGGT 658

RESULT 2
LOCUS M60982 459 bp mRNA EST 07-JUN-1996
DEFINITION zc98h09.s1 Pancreatic Islet Homo sapiens cDNA clone 339233 3'
ACCESSION M60982
NID g1367741
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondria; eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo;
1 (bases 1 to 459)
Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M.,
Holtman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maier, M.,
Patterson, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
TITLE Washu-Merck EST Project
JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: Exprimer
High quality sequence stop: 226.
Location/Qualifiers
1. .459
/organism="Homo sapiens"
/note="Organ: pancreas; Vector: pBluescript SK-; Site: 1;
EcORI; Site: 2; XhoI; Reference: Hum Mol Gen 2, 1795 (1993)
Takeda et al. Cloned unidirectionally. Primer: Oligo dT
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGACTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3' "
/db_xref="taxon:9606"
/db_xref="339233"
/clone="lib="pancreatic islet"
/issue="pancreatic islet"
/lab_host="SOLR cells (kanamycin resistant)"
complement(41..>459)

BASE COUNT 121 a 147 c 82 g 108 t 1 others
ORIGIN

Query Match 36.6% Score 450; DB 14; Length 459;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 454; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 TTTTATTTTATTTGACATTAATCAATTTTATTTGTTTCACAGAACATAGGCGATCT 60
|||||
Cp 1216 TTTTATTTTATTTGACATTAATCAATTTTATTTGTTTCACAGAACATAGGCGATCT 1157
Db 61 CGACAGTCGCTCCGTGACAGCCCAACCCCAACCCCTTACTGCGACCCACCTTAA 120
|||||
Cp 1156 CGACAGTCGCTCCGTGACAGCCCAACCCCAACCCCTTACTGCGACCCACCTTAA 1097
Db 121 GGGCAGTTCAAGAGAGAGAGAGATCTCAGAGATCTCATTCTAATGATGCGCGAGAGTC 180
|||||
Cp 1096 GGGCAGTTCAAGAGAGAGAGATCTCAGAGATCTCATTCTAATGATGCGCGAGAGTC 1037
Db 181 TCACACAGTAGACAGAGAGAGATGCTGAGATGAGATGAGATGAGATGAGATGAGATGAG 240
|||||
Cp 1036 TCACACAGTAGACAGAGAGAGATGCTGAGATGAGATGAGATGAGATGAGATGAGATGAG 977
Db 241 ACCCAGCAGAGACTTATCCAGCCGAGAGAGTCTCCACCCAGAGTCTCCCATTTT 300
|||||
Cp 976 ACCCAGCAGAGACTTATCCAGCCGAGAGAGTCTCCACCCAGAGTCTCCCATTTT 917
Db 301 CTCTCCTACTTTGCGGAGTTCAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
|||||
Cp 916 CTCTCCTACTTTGCGGAGTTCAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857
Db 361 GAATACCAAGAGAGCTCATTTTACAGCAGGAGGAGACATCTCACACCCCTTGNATTAAGTTAA 420
|||||
Cp 856 AATACCAAGAGAGCTCATTTTACAGCAGGAGGAGACATCTCACACCCCTTGNATTAAGTTAA 797
Db 421 AATACATATTAAGTACATCTCCATCCCTAGAGAGAC 459
|||||
Cp 796 AATACATATTAAGTACATCTCCATCCCTAGAGAGAC 758

RESULT 3
LOCUS AA047094 470 bp mRNA EST 06-SEP-1996
DEFINITION zK74602.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone
488522 5'
ACCESSION AA047094
NID g1525011
KEYWORDS EST.

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
9	W52269	413 bp mRNA	EST	11-OCT-1996								
	z464610.s1	Soares senescent fibroblasts	NBHSF	Homo sapiens	CDNA							
	clone 325338 3'											
	W52269											
	g1349381											
	EST.											
	EST.											
	human.											
	Homo sapiens											
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 413)											
	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kuabada, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.											
	The Washu-Merck EST Project											
	Unpublished (1995)											
	Contact: Wilson RK											
	Washu-Merck EST Project											
	Washington University School of Medicine											
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108											
	711: 314 286 1800											
	Fax: 314 286 1810											
	Email: est@wustl.edu											
	This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.											
	Insert Length: 3171 Std Error: 0.00											
	Seq primer: mob.RECA+ET.											
	Location/Qualifiers											
	1. 413											
	/organism="Homo sapiens"											
	/note="Vector: pT7SD (Pharmacia) with a modified polylinker V-type; phagemid; Site_1: Not I; Site_2: EcoRI; tetracycline-resistance gene; lacZ gene; lacI gene; lacO gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene; lacD gene; lacE gene; lacF gene; lacG gene; lacH gene; lacI gene; lacJ gene; lacK gene; lacL gene; lacM gene; lacN gene; lacO gene; lacP gene; lacQ gene; lacR gene; lacS gene; lacT gene; lacU gene; lacV gene; lacW gene; lacX gene; lacY gene; lacZ gene; lacA gene; lacB gene; lacC gene											

3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo.

/db.xref="taxon:9606"
/clone="325338"
/tissue_type="senescent fibroblasts NbHSF"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>413)

BASE COUNT 110 a 136 c 77 g 89 t 1 others
ORIGIN

Query Match 31.2%; Score 383; DB 15; Length 413;
Best Local Similarity 98.8%; Pred. No. 0.00e+00;

Matches 408; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

1 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGGCGATCTGACAGTGCCTC 60
|||||
CP 1204 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGGCGATCTGACAGTGCCTC 1145
DB 61 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 120
|||||
CP 1144 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 1085
DB 121 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGTCTTCACACAGTGA 180
|||||
CP 1084 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGTCTTCACACAGTGA 1025
DB 181 CAGACGAGTTNAGATGTTGATGAGATGACAGTCACTCTCAATCAAGCCACACACA 240
|||||
CP 1024 CAGACGAGTTNAGATGTTGATGAGATGACAGTCACTCTCAATCAAGCCACACACA 966
DB 241 GACTTCATCCAGCCGAGAGTCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 300
|||||
CP 965 GACTTCATCCAGCCGAGAGTCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 906
DB 301 TTTCGCGAGTTCAGAGTGTCTGCTTCCACACAGTCCACAAAGTCAATTAATTCACA 360
|||||
CP 905 TT-GCCGAGTTCAGAGTGTCTGCTTCCACACAGTCCACAAAGTCAATTAAT-ACCA 848
DB 361 GAGACCTGATTTACAGACAGGAGAACATCTCACACCTTGATTAAGTTAA 413
|||||
CP 847 GAGACCTGATTTACAGACAGGAGAACATCTCACACCTTG-CAATAAGTTAA 796

RESULT 10 N29782 390 bp mRNA EST 05-JAN-1996
LOCUS N29782.1 Homo sapiens cDNA clone 259649 3'
DEFINITION N29782.1 Homo sapiens cDNA clone 259649 3'
ACCESSION N29782
NID 91148302
KEYWORDS EST.
SOURCE human clone-259649 primer-m13 -40 forward library-Soares placenta 8to9weeks 2Nbp8to9w vector-pT73D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) Rstel-Not I Rstel2-Eco RI two placentae: one from 8 weeks and another from 9 weeks post conception. 1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGAGGAGGAGGCGGCGATTTTCTTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo.

ORGANISM

Homo sapiens

Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 390)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Location/Qualifiers

FEATURES
source 1..390
/organism="Homo sapiens"
/clone="259649"

BASE COUNT 103 a 129 c 75 g 83 t
ORIGIN

Query Match 30.6%; Score 376; DB 12; Length 390;
Best Local Similarity 98.7%; Pred. No. 0.00e+00;

Matches 381; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

5 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGGCGATCTGACAGTGCCTC 64
|||||
CP 1204 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGGCGATCTGACAGTGCCTC 1145
DB 65 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 124
|||||
CP 1144 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 1085
DB 125 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGTCTTCACACAGTGA 184
|||||
CP 1084 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGTCTTCACACAGTGA 1025
DB 185 CAGACGAGTTNAGATGTTGATGAGATGACAGTCACTCTCAATCAAGCCACACACA 244
|||||
CP 1024 CAGACGAGTTNAGATGTTGATGAGATGACAGTCACTCTCAATCAAGCCACACACA 966
DB 245 ACTTCATCCAGCCGAGAGTCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 304
|||||
CP 964 ACTTCATCCAGCCGAGAGTCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 905
DB 305 TGCCGAGTTCAGAGTGTCTCTCTCCACAGTCCACAAAGCTCAATTAATACCAAG 364
|||||
CP 904 TGCCGAGTTCAGAGTGTCTCTCTCTCCACAGTCCACAAAGCTCAATTAATACCAAG 845
DB 365 ACCTGCATTTACAGCAGGAGGAACAT 390
|||||
CP 844 ACCTGCATTTACAGCAGGAGGAACAT 819

RESULT 11 W47587 443 bp mRNA EST 11-OCT-1996
LOCUS W47587.1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone 32467 5' similar to SW:GI96_MOUSE P46594 IMMEDIATE EARLY PROTEIN G1Y96. [1] .
DEFINITION W47587.1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone 32467 5' similar to SW:GI96_MOUSE P46594 IMMEDIATE EARLY PROTEIN G1Y96. [1] .
ACCESSION W47587
NID 91332247
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryotes; Eukaryotes; Eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 443)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

CP 1024 CAGACGGAGTTGAGATGCTGGAGATGAGTACCTCTCTAAACTACGACCCACCCAG 965
 DB 244 ACTTCATCCACGCGGAGAGTCTCTCCCGACCCGAGTCTCTCCCACTTTCTCTACTT 303
 CP 964 ACTTCATCCACGCGGAGAGTCTCTCCCGACCCGAGTCTCTCCCACTTTCTCTACTT 905
 DB 304 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCACAGTCCACAAAGCTCAATTAATACCAAG 363
 CP 904 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCACAGTCCACAAAGCTCAATTAATACCAAG 845
 DB 364 ACCTGATTTA 374
 CP 844 ACCTGATTTA 834

RESULT 13
 LOCUS AA047184 355 bp mRNA EST 06-SEP-1996
 DEFINITION ZK4402.s1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone 48522.3
 ACCESSION AA047184
 NID 91525150
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryota; Eukaryota; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 355)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 324.
 FEATURES
 location/Qualifiers
 1..355
 /organism="Homo sapiens"
 /note="Organ: uterus; Vector: p773-Pac; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', AACGGAAGATTCGCGCGCGCTTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
 /db_xref="taxon:9606"
 /clone="48522"
 /clone_lib="Soares pregnant uterus NBHPU"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 complement(1..>355)
 BASE COUNT 91 a 121 c 66 g 76 t 1 others
 ORIGIN
 mRNA
 Query Match 28.6%; Score 351; DB 14; Length 355;
 Best Local Similarity 99.7%; Pred. No. 0.00e+00;
 Matches 351; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CP 1204 GACAGTAATCAATTTATTGTGTACAGAAATCTAGGAGATCTGACAGTGCCTC 1145
 DB 64 CGTAGAGCCACCAACCCCAACCTCTACCTCGAGCCACCTTAAGGAGACTTCAAG 123
 CP 1144 CGTAGAGCCACCAACCCCAACCTCTACCTCGAGCCACCTTAAGGAGACTTCAAG 1085
 DB 124 AAGATGGAAGATCTCAGGATCTCTATCTATATGTCGCCCGAAGTCTCAGACAGTGA 183
 CP 1084 AAGATGGAAGATCTCAGGATCTCTATCTATATGTCGCCCGAAGTCTCAGACAGTGA 1025
 DB 184 CAGACGAGTTGAGATGCTGGAGATGACAGTACCTCTTAATCTAGACCACACACAG 243
 CP 1024 CAGACGAGTTGAGATGCTGGAGATGACAGTACCTCTTAATCTAGACCACACACAG 965
 DB 244 ACTTCATCCACGCGGAGAGTCTCTCCCGACCCGAGTCTCTCCCACTTTCTCTACTT 303
 CP 964 ACTTCATCCACGCGGAGAGTCTCTCCCGACCCGAGTCTCTCCCACTTTCTCTACTT 905
 DB 304 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCACAGTCCACAAAGCTCAATTAAT 355
 CP 904 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCACAGTCCACAAAGCTCAATTAAT 853

RESULT 14
 LOCUS T49532 405 bp mRNA EST 08-FEB-1995
 DEFINITION ya76f12.s1 Homo sapiens cDNA clone 67631.3
 ACCESSION T49532
 NID 9651392
 KEYWORDS EST.
 SOURCE human clone=67631 library=Stratagene placenta (#937225)
 vector=pluscript SK- host=SOLR cells (kanamycin resistant)
 primer=21m3 Raitel-EcoRI Raitel2-XhoI Placental tissue from a Caucasian male. Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.2 kb; Uni-ZAP XR Vector. 5' adaptor sequence: 5'-GAATTCGGCAGCAG-3'; 3' adaptor sequence: 5'-CTCGAGTCTTTTCTTTTCTTTT-3'.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryota; Eukaryota; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Other ESTs: ya76f12.r1
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 FEATURES
 location/Qualifiers
 1..405
 /organism="Homo sapiens"
 /clone="67631"
 BASE COUNT 108 a 132 c 81 g 82 t 2 others
 ORIGIN
 Query Match 28.1%; Score 345; DB 9; Length 405;
 Best Local Similarity 96.3%; Pred. No. 0.00e+00;
 Matches 388; Conservative 0; Mismatches 9; Indels 6; Gaps 6;

Db 61 CAACCTTACCTCGCAGCCACCTAAGCGGACTTCAAGAAGATGGAAGATCTCAGG 120
 |||||||
 Cp 1124 CAACCTTACCTCGCAGCCACCTAAGCGGACTTCAAGAAGATGGAAGATCTCAGG 1065
 |||||||
 Db 121 ATCTCATCTAATAGTCCGCGGAAGTCTCACACAGTAGACAGAGAGTGGATGCTG 180
 |||||||
 Cp 1064 ATCTCATCTAATAGTCCGCGGAAGTCTCACACAGTAGACAGAGAGTGGATGCTG 1005
 |||||||
 Db 181 GAGATGAGTACCTCTCTAAGTACGACCCAGCAGACTTCATCCAGCGGAGC 240
 |||||||
 Cp 1004 GAGATGAGTACCTCTCTAAGTACGACCCAGCAGACTTCATCCAGCGGAGC 945
 |||||||
 Db 241 TCCCTCCCGCCAGCGAGTCTCCCATTTCTCTCTCTATCTGGCGGATTCAGGGGCTC 300
 |||||||
 Cp 944 TCCCTCCCGCCAGCGAGTCTCCCATTTCTCTCTCTATCTGGCGGATTCAGGGGCTC 886
 |||||||
 Db 301 CTGCTTCACACAGTCCCAAGAGTCAATTAATACAGAGAGAGAGAGTTCATTCAGGCA 360
 |||||||
 Cp 885 CTGCTTCACACAGTCCCAAGAGTCAATTAATACAGAGAGAGAGTTCATTCAGGCA 829
 |||||||
 Db 361 GAGGGAACATNTNACACCTTGGCATAGGTTTAAATAAT 403
 |||||||
 828 GAGGGAACATNTNACACCTTGGCATAGGTTTAAATAAT 788

RESULT 15 AA301628 413 bp mRNA EST 18-Apr-1997
 LOCUS EST14813 Aorta endothelial cells Homo sapiens cDNA 5' end.
 DEFINITION AA301628
 ACCESSION 91954112
 NID
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;

REFERENCE
 AUTHORS
 1 (bases 1 to 413)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wei,C.,
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geopaghen,N.S.,
 Glodex,A., Gnehm,C.L., Hanna,M.C., Hedlorn,E., Hinkle,P.S.,Jr.,
 Kelley,J.M., Kelley,J.C., Liu,L.-T., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudes,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Utecherback,T.R., Weidman,J.F., Li,Y.,
 Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dinke,D., Feng,D.-F., Ferrle,A., Fischer,C., Hastings,G.A.,
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
 Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
 Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
 Fraser,C.M. and Venter,J.C.
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)
 96026280
 JOURNAL
 MEDLINE
 COMMENT
 Other ESTs: TH015266
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tcbl/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES

SOURCE
 1. 413
 /organism="Homo sapiens"

/note="Organ: aorta; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI"
 /db_xref="ATCC (inhost):113941"
 /db_xref="taxon:9606"
 /clone_lib="Aorta endothelial cells"
 /cell_type="endothelial cell"
 /dev_stage="adult"
 <1..>413
 BASE COUNT 105 a 108 c 99 g 95 t 6 others
 ORIGIN
 mRNA
 Query Match 27.0%; Score 332; DB 18; Length 413;
 Best Local Similarity 96.8%; Pred. No. 0.00e+00;
 Matches 364; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

Db 1 CTGACCTCAGCACTTCTCTCCAGCAACACCGCGCCCTTCACTGTAATCCCGCA 60
 |||||||
 Qy 427 CTGACCTCAGCACTTCTCTCCAGCAACACCGCGCCCTTCACTGTAATCCCGCA 486
 |||||||
 Db 61 CTCGCCAAAAAAGATCCGAAAAACACAAAGAAACACAGCGCTGCTGCGGAGAG 120
 |||||||
 Qy 487 CTCGCCAAAAAAGATCCGAAAAACACAAAGAAACACAGCGCTGCTGCGGAGAG 546
 |||||||
 Db 121 N-TATCCCACTGAGGAGTCTCCGAGGCACTTGAACCTCAGAACACTACAGCGAGAGCC 179
 |||||||
 Qy 547 CGTATCCCACTGAGGAGTCTCCGAGGCACTTGAACCTCAGAACACTACAGCGAGAGCC 606
 |||||||
 Db 180 ACCCGGTGCTGAGGCGGAGCGAGCGGCGACAGAGACCGGCGCATAGAGACCGAGCA 239
 |||||||
 Qy 607 ACCCGGTGCTGAGGCGGAGCGGAGCGGCGACAGAGACCGGCGCATAGAGACCGAGCA 666
 |||||||
 Db 240 CAGCCAGCTGGGGGCTAGGCGCGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 299
 |||||||
 Qy 667 CAGCCAGCTGGGGGCTAGGCGCGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725
 |||||||
 Db 300 CTNCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 359
 |||||||
 Qy 726 CTCTT-AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 782
 |||||||
 Db 360 ACCTAATATTTATTT 375
 |||||||
 Qy 783 ACCTAATATTTATTT 798

Search completed: Thu Apr 23 05:39:21 1998
 Job time: 811 secs.

GN CEBPD OR CEF.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUKARYOTA; ROBOBETA.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 91332085.
RA KAGEYAMA R., SASAI Y., NAKANISHI S.;
RL J. BIOL. CHEM. 266:15525-15531(1991).
CC -1- FUNCTION: C/EBP IS A DNA-BINDING PROTEIN THAT RECOGNIZES TWO
DIFFERENT MOTIFS: THE CCAAT HOMOLOGY COMMON TO MANY PROMOTERS AND
THE ENHANCED CORE HOMOLOGY COMMON TO MANY ENHANCERS.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UNBIOLOGICALLY EXPRESSED.
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG. TO OTHER C/EBP
PROTEINS.
DR EMBL: M65149; G203412; -.
DR PIR: B39429; B39429.
DR TRANSFAC: T00109; -.
DR TRANSCRIPTION REGULATION: ACTIVATOR: DNA-BINDING: NUCLEAR PROTEIN.
DOMAIN 197 215 BASIC MOTIF.
DOMAIN 226 254 LEUCINE-ZIPPER.
SQ SEQUENCE 268 AA; 28600 MW; 760E79C0 CRC32;
Query Match 9.1%; Score 101; DB 1; Length 268;
Best Local Similarity 36.7%; Pred. No. 2.57e-01;
Matches 22; Conservative 10; Mismatches 25; Indels 3; Gaps 3;
Db 142 CAQTVSLAAAPPTSPSPSPSPSLAP-GPVREKAGKGRPPGSPPEYRQRER 200
Oy 7 CHTPTIIQAPT-PAPSTIPGRGSG-GPEITFDPLPEPAAPAGRSASGCHKRSR 64
RESULT 9
ID A1AB-CAEFA STANDARD: PRT: 417 AA.
AC P1615;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ALPHA-1B ADRENERGIC RECEPTOR (FRAGMENT).
GN ADRA1B OR RDC5.
OS CANIS FAMILIARIS (DOG).
CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUKARYOTA; CARNIVORA.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-THYROID;
RX MEDLINE: 89242119.
RA LIBERT F., PARMENTIER M., LEFORT A., DINSART C., VAN SANDE J.,
MAENHAUT C., SIMONS M.-J., DUMONT J.E., VASSART G.;
RL SCIENCE 244:569-572(1989).
CC -1- FUNCTION: THIS ALPHA-ADRENERGIC RECEPTOR MEDIATES ITS ACTION BY
ASSOCIATION WITH G PROTEINS THAT ACTIVATE A PHOSPHATIDYLINOSITOL-
CALCIN SECOND MESSENGER SYSTEM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL: X14050; G833784; -.
DR PIR: E30341; G30341.
DR GCRDB: GCR-0017; -.
DR PROSITE: PS00237; G-PROTEIN RECEPTOR. 1.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT NON-TER 1 1
FT TRANSMEM 1 8
FT DOMAIN 9 20
FT TRANSMEM 21 44
FT DOMAIN 45 64
FT TRANSMEM 65 85
FT DOMAIN 86 104
FT TRANSMEM 105 127
FT DOMAIN 128 198
EXTRACELLULAR (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 199 222
FT DOMAIN 223 229
FT TRANSMEM 230 243
FT DOMAIN 244 417
FT DOMAIN 271 281
FT DISULFID 21 98
FT LIPID 268 268
SQ SEQUENCE 417 AA; 45891 MW; 8008E9B3 CRC32;
Query Match 9.1%; Score 101; DB 1; Length 417;
Best Local Similarity 43.6%; Pred. No. 2.57e-01;
Matches 17; Conservative 10; Mismatches 10; Indels 2; Gaps 2;
Db 353 ALSLPAPQPGRRG-RDSCP-LFTFRLAERSSPAG 389
Oy 12 TILQPTPAPTICGPRGSGPEITFDPLPEPAAPAG 50
RESULT 10
ID QALF-NEUCR STANDARD: PRT: 816 AA.
AC P1638;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE QUINIC ACID UTILIZATION ACTIVATOR.
GN QA-1F.
OS NEUROSPORA CRASSA.
CC EUKARYOTA; FUNGI; ASCOMYCOTINA; PYRENOMYCETES; SORDARIALES.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-74-OR23-1A;
RX MEDLINE: 89293848.
RA GEEVER R.F., HUET L., BAUM J.A., TYLER B.M., PATEL V.B.,
RUTLEDGE B.J., CASE M.E., GILES N.H.;
RL J. MOL. BIOL. 207:15-34(1989).
CC -1- FUNCTION: TRANSCRIPTION ACTIVATION OF GENES FOR ENZYMES AND
PROTEINS OF QUINATE METABOLISM BY BINDING TO A 16 BASE-PAIR
SEQUENCE (CONSENSUS GGTARARYTTAYCC) IN FRONT OF EACH QA GENE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CLUSTER DOMAIN.
DR EMBL: X14603; G3068; -.
DR PIR: S04256; S04256.
DR PIR: F31277; F31277.
DR TRANSFAC: T00709; -.
DR PROSITE: PS00463; ZN2-CYS6-FUNGAL-1; 1.
DR PROSITE: PS00463; ZN2-CYS6-FUNGAL-2; 1.
KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; NUCLEAR PROTEIN;
KW ZINC; METAL-BINDING; QUINATE METABOLISM.
FT DNA_BIND 76 103
FT DOMAIN 773 793
SQ SEQUENCE 816 AA; 88946 MW; E59EACAT CRC32;
Query Match 9.1%; Score 101; DB 1; Length 816;
Best Local Similarity 33.3%; Pred. No. 2.57e-01;
Matches 16; Conservative 13; Mismatches 18; Indels 1; Gaps 1;
Db 408 VVRQLPEPDDRRPRRALLVACFYDVTYSRHHNPALKPDQD 455
Oy 70 VVRQLPEPDDRRPRRALLVACFYDVTYSRHHNPALKPDQD 116
RESULT 11
ID GSPC-ERNGH STANDARD: PRT: 272 AA.
AC P31696;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GENERAL SECRETION PATHWAY PROTEIN C (PEPTIC ENZYMES SECRETION PROTEIN
OUTC).
GN OUTC.
OS ERWINIA CHRYSANTHEMI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;

RC	TISSUE-EMBRYO;	
RX	MEDLINE; 92112033.	
RA	GRUBEY-JACKSON L., KUO A., LEDER P.;	
RL	GENES DEV. 6:29-37(1992).	
CC	-1- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB AND MAY HAVE A	
CC	FUNCTION IN DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING	
CC	SPECIFIC DIFFERENTIATED STATES	
CC	-1- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT	
CC	HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.	
CC	-1- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL	
CC	ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING	
CC	LIMB BUD.	
CC	-1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE	
CC	PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN	
CC	SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE	
CC	TRANSCRIPTS EXISTS WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS	
CC	DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR	
CC	DETERMINED.	
DR	EMBL; X62379; G51553; -	
DR	PIR; S24407; S24407.	
DR	HSSP; P19999; 1CLG.	
KW	NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.	
ET	DOMAIN 635 638 POLY-SER.	
FT	DOMAIN 644 744 PRO-RICH.	

RESULT	8	STANDARD:	PRT:	268 AA.
ID	CEBD_RAT			
AC	003484;			
DT	01-OCT-1996 (REL. 3.4, CREATED)			
DT	01-OCT-1996 (REL. 3.4, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 3.5, LAST ANNOTATION UPDATE)			
DE	CCAA/ENHANCER BINDING PROTEIN DELTA (C/EBP DELTA FACTOR CELL).			(TRANSCRIPTION

MUSE (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_p protein - protein database search, using Smith-Waterman algorithm
on: Fri Apr 17 13:38:32 1998; MasPar time 6.24 Seconds
626.917 Million cell updates/sec
Tabular output not generated.

Title: >US-08-799-910-10
Description: (1-156) from US08799910.pep
Perfect Score: 1114
Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPSPYALDLSTFLOQHPAAF 156

Scoring table:
PAM 150
Gap 11

Searched: 69112 segs, 25083644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-pro35
1:swiss1

Statistics: Mean 44.412; Variance 91.621; scale 0.485

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

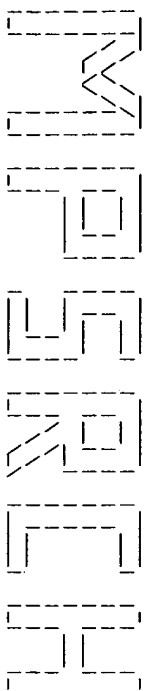
SUMMARIES

Alt	No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	1102	98.9	156	1	1	1EX1_HUMAN	RADIATION-INDUCIBLE IM	5.97e-191
2	752	67.5	153	1	1	1EX1_MOUSE	RADIATION-INDUCIBLE IM	1.15e-120
3	117	10.5	474	1	1	1VTP3_TTV1V	VIRAL PROTEIN TYP	2.07e-03
4	112	10.1	228	1	1	1VCOM_ADEMI	MINOR CORE PROTEIN (PR	9.76e-03
5	109	9.8	1799	1	1	1VO25_CAREL	HYPOTHETICAL 202.6 KD	2.43e-02
6	103	9.2	1206	1	1	1FOR4_MOUSE	FORMIN 4 (LIMB DEFORMI	1.44e-01
7	101	9.2	1468	1	1	1FORM_MOUSE	FORMIN (LIMB DEFORMITY	1.44e-01
8	103	9.1	268	1	1	1CEBD_RAT	CCAAT/ENHANCER BINDING	2.57e-01
9	101	9.1	417	1	1	1A1AB_CANFA	ALPHA-1B ADRENERGIC RE	2.57e-01
10	101	9.1	816	1	1	1QALE_NEUCR	QUINIC ACID UTILIZATIO	4.56e-01
11	99	8.9	272	1	1	1GSPC_ERWCH	GENERAL SECRETION PATH	6.05e-01
12	98	8.8	272	1	1	1GSQC_ERWCH	GENERAL SECRETION PATH	6.05e-01
13	98	8.8	390	1	1	1VGLI_HSV11	GLYCOPROTEIN 1	6.05e-01
14	98	8.8	475	1	1	1GAPN_STBMU	NADP-DEPENDENT GLYCERA	6.05e-01
15	98	8.8	519	1	1	1A1AB_HUMAN	ALPHA-1B ADRENERGIC RE	6.05e-01
16	98	8.8	975	1	1	1GDP_CANFA	CCAAT DISPLACEMENT PRO	6.05e-01
17	97	8.7	1043	1	1	1CHS2_PABR	CHITIN SYNTHASE 2 (EC	8.02e-01
18	97	8.7	1239	1	1	1V120_EBV	CAPSID ASSEMBLY PROTEI	8.02e-01
19	97	8.7	1337	1	1	1PTPJ_HUMAN	PROTEIN-TYROSINE PHOSP	8.02e-01
20	96	8.6	245	1	1	1VE4_HPVS5	PROBABLE E4 PROTEIN.	1.06e+00
21	96	8.6	245	1	1	1VE4_HPVS5	PROBABLE E4 PROTEIN.	1.06e+00
22	96	8.6	1233	1	1	1GUNA_HUMAN	GLUTAMATE (NMDA) RECEP	1.40e+00
23	95	8.5	234	1	1	1GUNA_HUMAN	GLUTAMINE SYNTHETASE (1.40e+00

24	95	8.5	316	1	1	1CDNC_HUMAN	CYCLIN-DEPENDENT KINAS	1.40e+00
25	95	8.5	449	1	1	1APG_BRANA	ANTER-SPECIFIC PROLINE	1.40e+00
26	95	8.5	493	1	1	1GUNA_XANCP	MAJOR EXTRACELLULAR EN	1.40e+00
27	95	8.5	534	1	1	1APG_ARATH	ANTER-SPECIFIC PROLINE	1.40e+00
28	95	8.5	643	1	1	1VP40_HSV2	CAPSID PROTEIN P40 (CO	1.40e+00
29	95	8.5	3865	1	1	1HRX_MOUSE	ZINC FINGER PROTEIN HR	1.40e+00
30	94	8.4	217	1	1	1YKRA_EBV	HYPOTHETICAL BKRF4 PRO	1.84e+00
31	94	8.4	219	1	1	1Y091_NFVOP	HYPOTHETICAL 29.3 KD P	1.84e+00
32	94	8.4	389	1	1	1NDPE_MOUSE	NPC DERIVED PROLINE RI	1.84e+00
33	94	8.4	1742	1	1	1GDP_CALSA	ENDOGLUCANASE A PRECUR	2.42e+00
34	93	8.3	232	1	1	1SAXI_CHICK	HOMEBOX PROTEIN SAX-1	2.42e+00
35	93	8.3	268	1	1	1CEBD_MOUSE	CCAAT/ENHANCER BINDING	2.42e+00
36	93	8.3	276	1	1	1YIT4_YEAST	HYPOTHETICAL 31.3 KD P	2.42e+00
37	93	8.3	350	1	1	1Y08N_MYCTU	HYPOTHETICAL 37.0 KD P	2.42e+00
38	93	8.3	464	1	1	1SP62_HUMAN	SPICEOSOME ASSOCIATED	2.42e+00
39	93	8.3	517	1	1	1MBEA_ECOLI	MOBILIZATION PROTEIN M	3.17e+00
40	92	8.3	620	1	1	1EXTN_TOBAC	EXTENSIN PRECURSOR (CE	3.17e+00
41	92	8.3	631	1	1	1DMK_MOUSE	MYOTONIN-PROTEIN KINAS	3.17e+00
42	92	8.3	820	1	1	1FIBL_HUMAN	FIBRINOGEN-LIKE PROTEI	3.17e+00
43	92	8.3	926	1	1	1YIKS_YEAST	HYPOTHETICAL 103.6 KD	3.17e+00
44	92	8.3	1692	1	1	1POLN_HBVPA	NON-STRUCTURAL POLYPRO	3.17e+00
45	91	8.2	440	1	1	1YWI9_MYCTU	HYPOTHETICAL 45.3 KD T	4.15e+00

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	156 AA.
ID	1EX1_HUMAN			
AC	P46595; Q93044;			
DT	01-NOV-1995 (REL. 32, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY			
GN	PROTEIN GLY96).			
OS	1EX1.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; PRIMATES.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-PLACENTA;			
RX	MEDLINE: 96181295.			
RA	KONDRATEYEV A.D., CHUNG K.N., JUNG M.O.;			
RL	CANCER RES. 56:1498-1502(1996).			
RN	(2)			
RP	PRELIMINARY SEQUENCE OF 1-106 FROM N.A.			
RC	TISSUE-PLACENTA;			
RA	HILLIER L., CLARK N., DUBOUQUE T., ELLISTON K., HAWKINS M., HOLMAN M.,			
RA	KIRKIN L., ROHLFING T., TAN F., TREVASKIS E., WATERSTON R.,			
RA	WILLIAMSON A., WOLDMANN P., WILSON R.;			
RL	SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- INDUCTION: BY RADIATION.			
CC	-1- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.			
CC	-1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO			
CC	FRAME-SHIFTS.			
DR	EMBL: S81914; G148385;			
DR	EMBL: T49531; -; NOT ANNOTATED_CDS.			
KW	GLYCOPROTEIN; TRANSMEMBRANE.			
FT	DOMAIN 1 82			
FT	TRANSMEM 83 99			
FT	DOMAIN 100 156			
FT	CARBOHYD 133 133			
SO	SEQUENCE 156 AA; 16973 MW; 380528CD CRC32;			
Query Match	98.9%;	Score 1102;	DB 1;	Length 156;
Best local Similarity	98.7%;	Pred. No. 5.97e-191;		
Matches 154;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;
DB	1 MCHSRCHPTMTILOAPTPA	STIPGRSGSPETFTDPLPEPAAPAGRSGSGRHK	60	
QY	1 MCHSRCHPTMTILOAPTPA	STIPGRSGSPETFTDPLPEPAAPAGRSGSGRHK	60	



(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

Mparch_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Apr 23 05:08:39 1998; Maspar time 999.58 Seconds

Tabular output not generated. 1452.649 Million cell updates/sec

Title: >US-08-799-910-9

Description: (1-1228) from US08799910.seq

Perfect Score: 1228 1 ATGTGTCACCTCTCGACGCTG.....AAAAAAAAAAAACTCGAG 1228

N.A. Sequence:

Comp: TACACAGTAGAGACGCTGAC.....TTTTTTTTTTTGGCTC

Scoring table: TABLE default

Gap 6

Match STD: Dbase 0; Query 0

Searched: 354530 seqs, 591221332 bases x 2

Post-processing: Minimum Match 0% Listing first 45 summaries

Database:

emb153

1:em_in 2:em_or 3:em_om 4:em_ov 5:em_pl 6:em_htg

7:em_hun1 8:em_hun2 9:em_ba 10:em_ro 11:em_un 12:em_v1

13:em_pat

genbank105

14:gb_ro 15:gb_om 16:gb_ov 17:gb_in 18:gb_pl 19:gb_da

20:gb_st 21:gb_v1 22:gb_ph 23:gb_sy 24:gb_un 25:gb_pat

26:gb_htg

Statistics: Mean 11.417; Variance 7.667; scale 1.489

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	118	9.6	1938	14	MMGLY96	1.49e-48
2	114	9.3	1758	14	RNPRL1	3.33e-46
3	79	6.4	7218	25	Sequence 14 from patent	4.27e-26
4	67	5.5	7218	25	Sequence 14 from patent	1.82e-19
5	41	3.0	215	25	Sequence 5 from patent	5.19e-06
6	37	3.0	10772	17	Drosophila melanogaster	4.03e-04
7	35	2.9	10772	17	Drosophila melanogaster	3.32e-03
8	33	2.7	215	25	Sequence 5 from patent	2.61e-02
9	29	2.4	800	16	Gallus gallus fibroblast	1.36e+00
10	29	2.4	1125	15	S. scrofa mRNA for L-36	1.36e+00
11	29	2.4	1266	15	BOVIOBP	1.36e+00
12	29	2.4	1294	14	DIROFASP70	1.36e+00
13	29	2.4	1479	14	Mus musculus mRNA for	1.36e+00
14	29	2.4	1811	16	Xenopus laevis nuclear	1.36e+00

C	15	29	2.4	1840	17	DDU73686	Dictyostelium discoide	1.36e+00
	16	29	2.4	2116	16	RCALN	Rana catesbeiana allan	1.36e+00
	17	29	2.4	3042	15	SS11545	S. scrofa mRNA for inte	1.36e+00
	18	29	2.4	3217	18	ZM85494	Zea mays LON1 protease	1.36e+00
	19	29	2.4	3737	14	S78355	Cyl-1-cyclin D1 (mice)	1.36e+00
C	20	30	2.4	3789	16	XELMYBRP1	Xenopus laevis myb-rel	5.19e-01
	21	28	2.3	571	17	SCU41441	Drosophila melanogaste	3.51e+00
	22	28	2.3	1404	16	S76875	casein kinase 2 alpha	3.51e+00
	23	28	2.3	1435	17	DMU75652	Drosophila melanogaste	3.51e+00
	24	28	2.3	1008	14	MMRNASB4	M. musculus seb4 mRNA..	3.51e+00
	25	28	2.3	2031	14	MMPTFRM	mouse mRNA for prothro	3.51e+00
	26	28	2.3	2179	25	E08204	cDNA encoding alpha 2,	3.51e+00
	27	28	2.3	2301	14	RNU75903	Rattus norvegicus UDP-	3.51e+00
	28	28	2.3	2624	17	AF020409	Dictyostelium discoide	3.51e+00
C	32	28	2.3	2628	25	166342	Sequence 1 from patent	3.51e+00
	33	28	2.3	289893	26	CEY47810	Dictyostelium discoide	3.51e+00
	34	27	2.2	479	14	BMORC5	Caenorhabditis elegans	3.51e+00
	35	27	2.2	714	18	BVNETVIB	Mus musculus nuclear c	8.87e+00
	36	27	2.2	861	14	AF025506	B. verrucosa Bet v 1b m	8.87e+00
	37	27	2.2	979	16	S7878954	Rattus norvegicus pren	8.87e+00
	38	27	2.2	1444	17	HELMAMIDE	CGAT-3 (chickens), liv	8.87e+00
	39	27	2.2	2240	17	LCU56636	H. echinata mRNA for LM	8.87e+00
	40	27	2.2	2498	14	MAU43333	Lucilia cupripa alpha	8.87e+00
	41	27	2.2	2842	25	A26375	Mesocricetus auratus 5	8.87e+00
	42	27	2.2	3016	17	DDU31631	RINCA coding sequence.	8.87e+00
	43	27	2.2	6258	17	PFSC03080	Dictyostelium discoide	8.87e+00
	44	27	2.2	8595	17	DDU14576	Plasmodium falciparum	8.87e+00
	45	27	2.2	27471	17	CEUC44C3	Dictyostelium discoide	8.87e+00
							Caenorhabditis elegans	8.87e+00

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM
1	MMGLY96	1938 bp	RNA	9287803	gly96 gene; glycosylated protein.	house mouse.	Mus musculus
							Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.
REFERENCE	AUTHORS	1 (bases 1 to 1938)					
	TITLE	Lau, L.					
	JOURNAL	Submitted (29-JUL-1992) L. Lau, University of Illinois College of Med., Dept. of Genetics, M/C 669, 808 South Wood Street, Chicago, IL 60612, USA					
REFERENCE	AUTHORS	2 (bases 1 to 1938)					
	TITLE	Charles, C. H., Yoon, J. K., Simske, J. S. and Lau, L. F.					
	JOURNAL	Genomic structure, cDNA sequence, and expression of gly96, a growth factor-inducible immediate-early gene encoding a short-lived glycosylated protein					
	MEDLINE	Oncogene 8 (3), 797-801 (1993)					
	FEATURES	93173526					
	source	Location/Qualifiers					
		1..1938					
		/organism="Mus musculus"					
		/strain="BALB/c"					
		/db_xref="taxon:10090"					
		BASE COUNT	390 a 525 c 503 g 520 t				
	ORIGIN						

Query Match Score 118; DB 14; Length 1938;

Best Local Similarity 77.8%; Pred. No. 1.49e-48;

Matches 165; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

DB 712 ATGTGCACTCGGCAACATCTCCACACCATGAGGCCCTTCTCCAGCT 771

QY 1 ATGTCTACTCTCGACGCTGCCACCCGACATGACCATCTCTCAGACCCCGACCCCGGCG 60
DB 772 CCCTCCACCGCGCGGAACTCCGGGGGCTGTGTCCGGAATTTGACCTTCGACCT 831
OY 61 CCTCCACCATCCCGGAGACCCCGGGGGCTCGGTCTGAGATCTTACCTTCGACCT 120
DB 832 CTCCTCGGAGCGGGCTGTGTCTCCACCGCGCTTGAACACTTCTCGGGGACCGAAG 891
OY 121 CTCCTCGGAGCGGGCTGTGTCTCCACCGCGCTTGAACACTTCTCGGGGACCGAAG 180
DB 892 CGCAGCCGAGGGTCTCTACCTTCGAGTGT 923
OY 181 CGCAGCCGAGGGTCTCTACCTTCGAGTGT 212

RESULT 2
LOCUS RNPRL 1758 bp DNA ROD 03-SEP-1996
DEFINITION R. norvegicus PRL1 gene.
ACCESSION X96437
NID g1515318
KEYWORDS PRL1 gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1758)
AUTHORS Schaefer, H., Trauzold, A., Siegel, E.G., Pilsch, U.R. and Schmidt, W.E.
TITLE PRL1: a novel early-response gene transcriptionally induced by
pituitary adenylate cyclase activating polypeptide in a pancreatic
carcinoma cell line
JOURNAL Cancer Res. 56 (11), 2641-2648 (1996)
MEDLINE 96221139
REFERENCE 2 (bases 1 to 1758)
AUTHORS Trauzold, A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular
Gastroenterology, 1st Dept. of Medicine, University of Kiel,
Schittenhelmstrasse 12, Kiel, D-24105, FRG

FEATURES
source
1. 1758
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/cell_line="AR4-2J"
590..1279
/gene="PRL1"
590..1279
gene /gene="PRL1"
BASE COUNT 358 a 475 c 470 g 455 t
ORIGIN

Query Match 9.3%; Score 114; DB 14; Length 1758;
Best Local Similarity 76.9%; Pred. No. 3.33e-46;
Matches 169; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

DB 590 ATGTCCATTGCGTACCACTTCACACCATGACTGCTGAGGCTCCGGCGGCACTT 649
OY 1 ATGTCTACTCTCGACGCTGCCACCCGACATGACCATCTCTCAGACCCCGGCG 60
DB 650 CCCTCCACCGCGCGGAACTCCGGGGGCTGTGTCCGGAATTTGACCTTCGACCT 709
OY 61 CCTCCACCATCCCGGAGACCCCGGGGGCTCGGTCTGAGATCTTACCTTCGACCT 120
DB 710 CTCCTCGGAGCGGGCTGTGTCTCCACCGCGCTTGAACACTTCTCGGGGACCGAAG 769
OY 121 CTCCTCGGAGCGGGCTGTGTCTCCACCGCGCTTGAACACTTCTCGGGGACCGAAG 180
DB 770 CGCAGCCGAGGGTCTCTACCTTCGAGTGT 801
OY 181 CGCAGCCGAGGGTCTCTACCTTCGAGTGT 212

RESULT 3

LOCUS 166494 7218 bp DNA PAT 23-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
NID g2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
1..7218
source /organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 5.5%; Score 67; DB 25; Length 7218;
Best Local Similarity 2.4%; Pred. No. 1.82e-19;
Matches 9; Conservative 214; Mismatches 156; Indels 0; Gaps 0;

DB 1061 TGCATATTT 1120
OY 99 TGAGATCTTCACCTTCGACCTCTCCGAGCCCGGAGCCCGCTGCGGGCGCCAG 158
DB 1121 YY 1180
OY 159 GCGCTTCGCGGCGACCGAAGCGACCGACCGAGTTCTTACCTCGAGTGTCCGGC 218
DB 1181 YY 1240
OY 219 CCAGCTCCGAGTCCGAGACCGAAGCGACCGAAGAGCTCTCTGTCTGCTACCAT 278
DB 1241 YY 1300
OY 279 CGTCTTCCGAGTCTCGATGAGTGAAGAGGTGTGCGCCCGCTGCTCGACAGGA 338
DB 1301 YY 1360
OY 339 GCGCCCTAAGCGCGCATCCCTGCGCGCCGACCCCTGTGCTCTCGAGCCCTTAA 398
DB 1361 YY 1420
OY 399 TCTGACTCGAGCGCTCGGAGTCTGAGCTCGAGCTTCTCTCGAGACACCC 458
DB 1421 YYYYYYYYYYYY 1431
OY 459 GCGCGCTTCT 469

RESULT 4
LOCUS 166494 7218 bp DNA PAT 23-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
NID g2724471
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
1..7218
source /organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN


```

source          1.215
                /organism="unknown"
BASE COUNT      15 a      8 c      25 g      26 t      141 others
ORIGIN
Query Match     2.74; Score 33; DB 25; Length 215;
Best Local Similarity 12.2%; Pred. No. 2.61e-02;
Matches 25; Conservative 79; Mismatches 101; Indels 0; Gaps 0;

Db 11 VVSRATSCNDKAKKDCNTSSMTTDCNRTMGVCDTDTTRVNVNDSGHNNSANYG 70
Oy 799 AACTTATCCAGAGGTGTGACATCTTCCCTGCTGTAATGACAGCTCTTGTATTTAT 858
Db 71 NNVAARHTHYTHTNVSGADSKTVDTSYNASGTSNGGTDGNSGADSYSSKTAATSR 130
Oy 859 TGAGCTTTGGCAGCTGTGAGAGACAGCACTGGAACTGGCGCAAGTAGAGAAGAA 918
Db 131 NRTGKTAANNAVDSRNMGDASVSDKNTTKAKNSADGKSKNNDRNNRYGTGTSNVS 190
Oy 919 ATGGGAGAGCACTGGGAGGAGAGCACTCCGCGTGGATGAAGTCTGCTGGGCTCG 978
Db 191 NNGGGNKRNVSSYANNKCGGSSCT 215
Oy 979 TAAGTTTAGGAGGTGACATCTCT 1003

RESULT 9
LOCUS          GGU1467      800 bp      mRNA      VRT      02-APR-1996
DEFINITION     Gallus gallus fibroblast growth factor 8 FGF8 mRNA, partial cds.
ACCESSION      U41467
NID            91134969
KEYWORDS
SOURCE         Chicken.
ORGANISM       Gallus gallus
                Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
                Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
                Phasianidae; Phasianinae; Gallus.
REFERENCE
AUTHORS        Crossley P.H., Minowada G., MacArthur C.A. and Martin G.R.
TITLE          Roles for FGF8 in the induction, initiation, and maintenance of
                chick limb development
JOURNAL        Cell 84 (1), 127-136 (1996)
MEDLINE
REFERENCE      2 (bases 1 to 800)
AUTHORS        Crossley P.H., Minowada G., MacArthur C.A. and Martin G.R.
TITLE          Direct Submission
JOURNAL        Submitted (27-NOV-1995) Philip H. Crossley, Anatomy, UCSF,
                Parnassus, San Francisco, CA 94143-045, USA
FEATURES
source
1.800
/organism="Gallus gallus"
/db_xref="taxon:9031"
/dev_stage="embryo"
/tissue_type="brain"
/clone_lib="E10 embryonic chick brain cDNA library"
<1.615
/codon_start=1
/product="fibroblast growth factor 8 FGF8"
/db_xref="PID:g1134970"
/translacion="VPMHLFVLCLOAQVTVSPNFTOHVREGSLVTDLSRLVRYT
OLYSRTSGHVOILDNKINAMADGDVAKLIVETDFTGSRIRKGAATGTFYCNK
KGLIKSGNKGKDVETEIVLENNYTAIONAYEENYMAFTKRGPRGSKTRQHOR
EVHFMKRLKXGHTTEPHRFELNYPFNRSKRRTNSSARVP"
BASE COUNT      225 a      219 c      226 g      130 t
ORIGIN
Query Match     2.4%; Score 29; DB 16; Length 800;
Best Local Similarity 74.6%; Pred. No. 1.36e+00;
Matches 44; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 742 TGTCTCTGTTTTACAAAAAACCAACAAACAAAAAAGAACTCGAG 800
Oy 1170 TGTCTCTGTAACAAATATGATTGATTCTGTCAAAAAAAGAAAGAACTCGAG 1228

```

```

RESULT 10
LOCUS          SSL36LBP      1125 bp      RNA      MAM      11-JAN-1995
DEFINITION     S.scrofa mRNA for L-36 lactose binding lectin.
ACCESSION      X79303
NID            9623345
KEYWORDS       1-36LBP gene; lactose-binding lectin.
SOURCE         Pig.
ORGANISM       Sus scrofa
                Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
                Vertebrata; Eutheria; Artiodactyla; Suidae; Sus.
REFERENCE
AUTHORS        Chiu M.L., Parry D.A., Feldman S.R., Klapper D.G. and O'Keefe E.J.
TITLE          An adherens junction protein is a member of the family of
                lactose-binding lectins
JOURNAL        J. Biol. Chem. 269 (50), 31770-31776 (1994)
MEDLINE
REFERENCE      2 (bases 1 to 1125)
AUTHORS        O'Keefe E.J.
TITLE          Direct Submission
JOURNAL        Submitted (18-MAY-1994) E.J. O'Keefe, University of North Carolina,
                137 NCMH, Dept of Dermatology, CB7600, Chapel Hill, NC 27514, USA
FEATURES
source
1.1125
/organism="Sus scrofa"
/db_xref="taxon:9823"
/dev_stage="adult"
/tissue_type="epithelium, oral, tongue"
/cell_type="keratinocyte"
/clone_lib="pig tongue cDNA"
49.1020
/gene="136LBP"
49.1020
/gene="136LBP"
/product="L-36 lactose binding protein"
/db_xref="PID:9623346"
/translacion="MAFVAPRGVPTVPTPYKPIPGSLRVGMSVYIIGVANEHMK
RFVVFVVGCGPAGADVAFHNPREDGMKVFYNSODCKMNEKREMPRAPAPE
LVNLYLPHRIVVYVNGDPFFYFGRHPIVQVTHLOVDDLTLOSINFITGGVPAPSPG
MNPDPYGPGRHNDQPCNLPCMEGAPTEPVPVYKTRKQGLVARITIVIGVPPSG
KSLVNFVSGSDVALINPLETIGIVRNSYLVNGKGADEKRSSEFPAFGVFDL
SIRGLDFRKYVANGOHLPFDSHRISNFGVDVTLIOGDVTLVSVQI"
BASE COUNT      255 a      347 c      303 g      220 t
ORIGIN
Query Match     2.4%; Score 29; DB 15; Length 1125;
Best Local Similarity 89.2%; Pred. No. 1.36e+00;
Matches 33; Conservative 0; Mismatches 4; Indels 5; Gaps 0;

Db 1089 AATAAATTAATTAATGCGCAAAAAAAGAAAAA 1125
Oy 1185 AATAAATTAATTAATGCGCAAAAAAAGAAAAA 1221

RESULT 11
LOCUS          BOV10PP      1266 bp      mRNA      MAM      25-JAN-1993
DEFINITION     Bovine inorganic pyrophosphatase mRNA sequence.
ACCESSION      M95283
NID            9163228
KEYWORDS       inorganic pyrophosphatase.
SOURCE         Bos taurus retina cDNA to mRNA.
ORGANISM       Bos taurus
                Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
                Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
                Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS        Yang Z. and Wensel T.G.
TITLE          Molecular cloning and functional expression of cDNA encoding a
                mammalian inorganic pyrophosphatase
JOURNAL        J. Biol. Chem. 267, 24641-24647 (1992)

```

FEATURES	Location/Qualifiers
source	1. .1266
	/organism="Bos taurus"
	/db_xref="taxon:9913"
	/tissue_type="retina"
BASE COUNT	396 a 239 c 300 g 331 t
ORIGIN	
Query Match	2.4%; Score 29; DB 15; Length 1266;
Best Local Similarity	87.2%; Pred. No. 1.36e+00;
Matches	34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db	1227 AATATAAATTTATTTGTTCAAAAAAAAAAAAAAAAAA 1265
Oy	1184 AATATAAATGATTTACTGTCTCAAAAAAAAAAAAAAAAAA 1222
RESULT 12	
LOCUS	DIRHSP70 1294 bp mRNA INV 25-AUG-1992
DEFINITION	Dirofilaria immitis 70 kDa heat shock protein homologue (hsp 70)
ACCESSION	U05648
NID	M95648
KEYWORDS	9156705
SOURCE	heat shock protein 70; heat shock protein homologue.
ORGANISM	Dirofilaria immitis adult cDNA to mRNA.
	Dirofilaria immitis
	Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda;
	Secernentea; Spirurina; Spirurida; Spirulina; Filarioidea;
	Onchocercidae; Dirofilaria.
	1 (bases 1 to 1294)
REFERENCE	Culpepper, J.A., Friedman, L. and Dale, B.
AUTHORS	Molecular cloning and characterization of a Dirofilaria immitis
TITLE	cDNA encoding an Hsp 70 homologue
FEATURES	Unpublished (1992)
source	Location/Qualifiers
	1. .1294
	/organism="Dirofilaria immitis"
	/db_xref="taxon:6287"
	/dev_stage="adult"
	1. .1038
	/partial
	/note="homologue"
	/codon_start=1
	/product="heat shock protein 70"
	/db_xref="PID:g156706"
	/translation="RFEELCDADLRSTMDPVEKALRDAMKKAQVHDIYVYGSTRIF
	KVKKLLSDPFGSKELINKSINPDEAVAGAAVQAAILISGKSEAYQDILLIDVAFSLG
	LETRGGVATLIRKNTTTPITSQTFITSONQGVLLQVIEGSRAMTKDNLLGKFE
	LSGIPPARRGVPOILEVTFIDANGLVNSADQSTGQNKTTINDGRSKDEIERM
	VOEAEKRYADDEAKDRIAAKNNAESYAFNNKQTIIEDEKLDKISEDKIKKIQKCDCE
	TVRLDNGNTAEKDEFEHROKELSEYCNPIITTKLYQASGMPGMPGMPGAPGAGGSS
	TGGGPTTIEEVD"
BASE COUNT	409 a 229 c 296 g 360 t
ORIGIN	
Query Match	2.4%; Score 29; DB 17; Length 1294;
Best Local Similarity	87.2%; Pred. No. 1.36e+00;
Matches	34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db	1256 AAGTAAATTTATATGCGCTCAAAAAAAAAAAAAAAAAA 1294
Oy	1184 AATATAAATGATTTACTGTCTCAAAAAAAAAAAAAAAAAA 1222
RESULT 13	
LOCUS	NM265PROT 1479 bp RNA ROD 07-OCT-1997
DEFINITION	Mus musculus mRNA for 26S proteasome non-ATPase subunit.
ACCESSION	Y13071
NID	G2505939
KEYWORDS	26S proteasome; non-ATPase subunit; proteolysis; regulatory
SOURCE	complex.
ORGANISM	house mouse.
	Mus musculus

REFERENCE
AUTHORS
TITLE

JOURNAL
REFERENCE
AUTHORS
TITLE

FEATURES
Source

gene
CDS

Unpublished
2 (bases 1 to 1479)
Gordon,C.
Direct Submission
Submitted (08-MAY-1997) C. Gordon, MRC Human Genetics Unit,
Developmental Genetics, Western General Hospital, Crewe Road,
Edinburgh EH4 2XU, UK
Location/Qualifiers
1..1479
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="adult"
237..1166
/gene="PADI"
237..1166
/gene="PADI"
/function="proteolysis"
/codon_start=1
/product="26S proteasome, non-ATPase subunit"
/db_xref="PID:e352083"
/db_xref="PID:g2505940"
/translation="MDRLRLRGMPGLARPDPDAPVADPAEDVYISLLALKLKHG
RAGPVNFMGLMGFEFDDTVIVIDFAFAMPGSTGVSAVPYPQAKHLDMLKQG
RPKVYGVNHSHFGFCWLSGVIDINTOOSFALSERAAVVADTSVKRVVDAR
LINNMNVLHEPRQTSLNGHLNPISQILGLNHRYSITINRNPLEDEKMILLN
LHKSNMEGLTLDYSECHKNSVSVMELAKNKYKAVEEEDKMTPEOLAIKNKG
ODPKRHLEHVDTLMNTNIYQCLAAAMDVTVF"

BASE COUNT 453 a 270 c 349 g 407 t

ORIGIN

Query Match 2.4%; Score 29; DB 14; Length 1479;
Best Local Similarity 87.2%; Pred. No. 1.36e+00;
Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1440 AACTTAATACATTTCACATCCAAAAA...AAAAAAAA 1478
||||||| ||||| || |||||
QY 1184 AATAAATTGATTACTCTCAAAAAA...AAAAAAAA 1222

RESULT 14
LOCUS XLU69669 1811 bp mRNA VRT 26-FEB-1997
DEFINITION Xenopus laevis nuclear pore complex-associated protein TPR (tpr)
ACCESSION U69669
MIM g1850343
KEYWORDS African clawed frog.
SOURCE
ORGANISM
Xenopus laevis
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
Pipidae; Xenopodinae; Xenopus.
1 (bases 1 to 1811)
Cordes,V.C., Reidenbach,S., Rackwitz,H.R. and Franke,W.W.
Identification of protein p270/Tpr as a constitutive component of
the nuclear pore complex-attached intranuclear filaments
J. Cell Biol. 136 (3), 515-529 (1997)
97177132
2 (bases 1 to 1811)
Cordes,V.C., Hunzicker,A. and Franke,W.W.
Direct Submission
Submitted (06-SEP-1996) Cell Biology/0110, German Cancer Research
Center, INF 280, Heidelberg 69120, Germany
Location/Qualifiers
1..1811
/organism="Xenopus laevis"
/db_xref="taxon:8355"

FEATURES
Source

[illegible]

CC 700 CTCCTCCACACGGGGCCATGACCCAGCTGGGGCTGGCTGGCTCTCATAGGCGCTCGGTC 64

Db 268 ratg-rnmvndrltrnnananrnanrnanrntvntlynnannnnnnnnnnrnnrnnrnnndgnyng 32

CP 640 TCTGTGGCCCTCGGGTCCCGCCCTCAAGACACCGGGTGGGCTCTCCGCTAGTGGTCTGAGT 581

Db 327 nsmnnnnaagcnydgynnnnyanvnnhntlmnggrindgnrvn-kmgrryhgltgryvndkn 38

CP 560 TCMAATTCCTCCGGAAGGCCACAGTTGGGGATACGCTCTCGCGCACAGTACGGCTGGTG 522

Db 386 ndrncdvrvwmgdngsdgnnaahysganrknwvtgrnnrvkxgamsndnnandnd 44

CP 520 TTTCTTTGTGGTTTTTTCGGATTCTTTTGGGGAGTGGCGGAGTCACAGTTAAGAAAGCGG 46

Db 446 nscdktnnstnanvanygtgnltmnyvssnnnnrkkmnnknaasmvrvrvnnnnngsn 50

CP 460 CCGGGTGTTCGTGCGAAGAAATGCTGACGTCACAGTCGATGTCGAGGCGCTCCGAATCA 40

Db 506 ryhkpa-gsrntnstrg-sysysmtalgykymnanbtknrvanbhvkkrnnnltln 56

CP 400 GATTAAAGGGCTCGAGGAGCGGGGACACAGGGGTGGCGCCAGGAGATGGCGCTTAGGGG 34

Db 564 vnnnnkhmrdrvnnhnttrngacdnnnnncvntvncrsgndnnndnnndnnmnyrn 62

CP 340 CGTCTCTTGGAGGCGGGGGCCCGCCACACCTCTTCACGCTACAGATTCGCGAGAAGA 28

Db 624 nndnvkymannhnsnshgnsksncvdrsnvknltldyngnastrnanddnanyak 68

CP 280 CGATGTGTGAGCA-GAAGAAGAACCTTTGGCTGGTTCGCTTC-TCGACTGGCAG 22

Db 684 knntannnsgnnnttymnaadvysngnnnnnnnrgnnymngndnknnykvng 74

CP 222 CTGGCGCGGAGCACTCAGAGGTA-GAAGAACCTCGCGCTCGCTTTGGTGGCGCGAG 16

Db 744 nrvnrvnrdtrnnnnnnhvmnmrvwandnarnrdngnknrxrnnkngtstndnnn 80

CP 163 AGCGCTCGGGCGCCCGGCGAGGGCGCTCGGGCTCCGGGAGAGGTCGAAGTGAAGA 10

Db 804 rnmnyannnkvnrtlnaynnkkrkananynnnhsvannnkrgntvnanandsvny 86

CP 103 TCTCAGAACCGAGCGCCCGCGGGTCCCGGATGTGGAAGGGGCGGGGTGGGGCT 44

Db 864 nsdvngtansanstnmvntnndnytcndandndv 902

CP 43 CGAGATGTCATGCTCGGGTGGCGAGCTGCGAGAGTAC 5

RESULT 4

ID 077554 standard; DNA: 267 BP.

AC 077554:

DT 23-SEP-1994 (first entry)

DE Human genome fragment. (Preferred)

KW Brain; placenta; bone marrow; genetic analysis; gene mapping; detection; homology; human; adrenal tissue; ds.

OS Homo sapiens.

PN WO9401548-A.

PD 20-JAN-1994.

PF 13-JUL-1993: G01467.

PR 13-JUL-1992: GB-014857.

PA (MEDIC) MEDICAL RES COUNCIL.

PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D; Pi Sibson DR, Starkey M; WPI: 94-035056/04.

PT New nucleic acid fragment encoding gene products - can be used for genetic analysis and mapping

PS Claim 1; Page 566; 616pp; English.

CC Human nucleic acid fragments, isolated from brain, adrenal tissue, the placenta or bone marrow comprise any of: (A) a sequence selected from (076401-077613), (B) an allelic variation of a sequence as described in (A), or (C) a sequence complementary to (A) or (B).

CC Preferred sequences exhibit no more than 90105mology to a human sequence known per se.


```

FT /tag=
FT /function=multple cloning site
FT primer_bind 187..204
FT
FT     /tag- b
PN EP-265123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034619.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
PT Introducing random point mutations into nucleic acids -
PT misincorporation, completion of molecules and screening.
PS Disclosure; p: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match      3.5%; Score 43; DB 1; Length 204;
Best Local Similarity 14.5%; Pred. No. 2,31e+08;
Matches 18; Conservative 61; Mismatches 44; Indels 1; Gaps 1

Dp   62 ywccgagcygcaaycdchvgcgcmrttthyrmphnyrvyrdmsdaawccyrsv 121
    :||| ||| |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp  1134 CACCAACCCCCAACACCCTTACTTGCGACGCCACCTTAAGGCGACTTCAGAAGA-TGGAA 1076
    :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db   122 kydcynacchdhdyvbbbyvnyhnbnnccecbnhvchnvbhnbhnmhvayrhdar 181
    :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp  1075 GGATGTCAACGATGCATCTCATTAATGCGCCGCCGAAGTCTCACACAGTAGACAGCGAG 1016
    :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db   182 ddtvh 185
    ::::

Cp  1015 TTGA 1012

RESULT          9
ID NB1164 standard; DNB: 204 BP.
AC NB1164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
TH Key Location/Qualifiers
FT misc_feature 19..69
FT     /tag- a
FT     /function=multple cloning site
FT     primer_bind 187..204
FT
FT     /tag- b
FN EP-265123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
PT WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded templates, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to

```

it to generate a popn of DNA molecules which terminate at all possible nucleotide positions within a specified region. The variable 3' ends generated in this way are used as primers for reverse transcriptase. Nucleotides are misincorporated by the transcripase and the molecules are completed to forms that can be amplified and then expressed in a suitable host-vector system. The sequence covers all 176 diff base substitutions, most of which occurred singularly in any given mutant.

See also P80575.

Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.3%; Score 41; DB 1; Length 204;
Best Local Similarity 18.7%; Pred No. 2,10e-07;
Matches 29; Conservative 69; Mismatches 54; Indels 3; Gaps 2;

Db 24 cyyrcayccbgagagycgacbcyrtagngyccccggggwccgcygaaycdchvcg 83
::: : ::::
OY 9 CTCCTCCAGCTGCACCCGCGCATGCATCCTGGACGCCGCCGCCGCCCTC-- 65

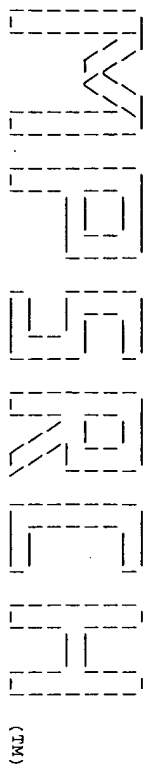
Db 84 cgyrmttthyrmrhnyrdynrsdaawcycrryxkydcgnachdhhyrbbyv 143
:
OY 66 CACCATCCCGGAGCCCCGGGGGGCGTCTTGATGTTCACCTTCGACCTTCGCC 125

Db 144 nvhhmnccebnhvchnvbhnhrwayrhd 178
:
OY 126 GGAGCCCGCAGCGGCCCTGCGGGGCCGCCGCCGCCGCCGCCGCCGCCGCC 160

RESULT 10
ID O70469 standard; DNA: 114 BP.
AC O70469;
DT 07-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
sequence of 6,9 or 12 nucleotides (see
FT comments)"
FN W05418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994: U00977.
PR 01-FEB-1993: US-013416.
PR 30-DEC-1993: US-176500.
PR 31-JAN-1994: US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
WP1: 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
screening a recombinant vector library expressing fusion proteins
comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC O70469 is a generic DNA sequence used to generate random TSAR peptide
library. The sequence is represented as follows: X(TGC)(NNB)10-(
TGC)(NNB)6(NNB)2(TGC)(NNB)14(TGC)x. x and y are flanking restriction
sites (x is not the same as y) that are not specified further. This
sequence generates peptides that are cloverleaf in structure. Other
generic sequences are shown in O70465-68. Other specific peptides
generated by these generic sequences are shown in R65150-54. TSARs are
concatenated heterofunctional proteins or peptides, comprising at least
two functional regions - a binding domain with affinity for a ligand and
a second effector peptide portion that is chemically or biologically
active. They may further comprise a linker peptide between the 2 domains.
The oligonucleotides are also designed so that the expressed peptide
contains 2 or 4 cysteine residues positioned in, or flanking, the
unpredicted or variant residues. These residues confer some degree of
conformational rigidity to the peptides. The TSARs or comps, comprising
a TSAR binding domain can be used in vivo to deliver a chemically or
biologically active moiety, eg. metal ion, radioisotope, peptide, toxin

RESULT 13
ID 070467 standard: DNA: 114 BP.
AC 070467;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /*tag= a
FT /note= "this sequence represents '2'; 2 can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR MPI: 94-279739/34.
DR P-PSDB: R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PS Disclosure: Page 35; 255pp; English.
CC 070467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)16(TGC)(NNB)1Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in 070466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Query Match 2.9%; Score 36; DB 12; Length 114;
Best Local Similarity 3.7%; Pred. No. 4.59e-05;
Matches 4; Conservative 33; Mismatches 71; Indels 0; Gaps 0;
Db 3 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnnnnnn 62
Cp 696 TTCCACCGGCGCTAGCCGCGCTGCTGCTCTCTATGCCCGCTGCTCTG 637
Db 63 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbg 110
Cp 636 TGGCGCTCGTCCGCTCAGACACCGGCTGCGCTCTCCGCTGAGTG 589
RESULT 14
ID 070466 standard: DNA: 114 BP.
AC 070466;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR-9 peptide library.

KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /*tag= a
FT /note= "this sequence represents '2'; 2 can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PF 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR MPI: 94-279739/34.
DR P-PSDB: R65152.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PS Disclosure: Page 35; 255pp; English.
CC 070466 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)1(TGC)(NNB)10(TGC)2(NNB)8(TGC)(NNB)
CC -9Y. X and Y are flanking restriction sites (X is not the same as Y)
CC that are not specified further. Other generic sequences are shown in
CC 070466-68. Other specific peptides generated by these generic sequences
CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC or peptides, comprising at least two functional regions - a binding
CC domain with affinity for a ligand and a second effector peptide portion
CC that is chemically or biologically active. They may further comprise a
CC linker peptide between the 2 domains. The oligonucleotides are also
CC designed so that the expressed peptide contains 2 or 4 cysteine residues
CC positioned in, or flanking, the unpredicted or variant residues. These
CC residues confer some degree of conformational rigidity to the peptides.
CC The TSARs or compns. comprising a TSAR binding domain can be used in
CC vivo to deliver a chemically or biologically active moiety, eg. metal
CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC on the cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;
Query Match 2.9%; Score 36; DB 12; Length 114;
Best Local Similarity 7.3%; Pred. No. 4.59e-05;
Matches 8; Conservative 31; Mismatches 71; Indels 0; Gaps 0;
Db 5 gcdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnn 64
Cp 703 GCTCTCTCTCCACCGGCGCTAGCCGCGCTGCTGCTGCTCTATGCCGCTCG 644
Db 65 nbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnnb 114
Cp 643 GTCCTGTGCGCTGCTGCCGCTCAGACACCGGCTGCGCTCTCCGCTG 594
RESULT 15
ID T76405 standard: DNA: 178 BP.
AC T76405;
DT 15-SEP-1997 (first entry)
DE Human endothelin-1 antisense oligonucleotide.
KW Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW chronic obstructive pulmonary disease; bronchitis; ss.
OS Synthetic.
PN WO9640162-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09306.
PR 07-JUN-1995; US-474497.



Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Apr 17 13:41:02 1998; Maspar time 6.99 Seconds
310.069 Million cell updates/sec
Tabular output not generated.

Title: >US-08-799-910-10
Description: (1-156) from US08799910.pep
Perfect Score: 1114
Sequence: 1 MCHSRSCHTMTILQAPTPA.....EPDVALDISTFIQHPAPF 156

Scoring table:
PAM 150
Gap 11

Searched: 111726 seqs, 13889129 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneseq30
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23

Statistics: Mean 31.062; Variance 153.160; scale 0.203

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	101	9.1	183	23	W14570	Streptococcus pneumoniae
2	100	9.0	439	9	R28150	Sugar beet chitinase
3	98	8.8	520	9	R53072	Alpha 1b adrenoceptor
4	98	8.8	520	9	R52831	Sequence of human alp
5	98	8.8	520	14	R85943	Alpha-1b adrenoceptor
6	97	8.7	190	23	W14569	Streptococcus pneumoniae
7	97	8.7	1337	12	R85203	hudep-1.
8	96	8.6	775	12	R66057	Human NMDAR2 receptor
9	96	8.6	1214	12	R66065	Human NMDAR2 receptor
10	96	8.6	1219	12	R66063	Human NMDAR2 receptor
11	96	8.6	1231	12	R66062	Human NMDAR2 receptor
12	96	8.6	1236	12	R66037	Human N-methyl-D-aspa
13	96	8.6	1239	12	R66064	Human NMDAR2 receptor
14	96	8.6	1244	12	R66061	Human NMDAR2 receptor
15	95	8.5	198	10	R59841	ApoE4L protease.
16	95	8.5	198	15	R92113	Human ApoE4L.
17	95	8.5	311	10	R59843	ApoE4Lx2 protease.
18	95	8.5	311	15	R92115	Human ApoE4Lx2.
19	94	8.4	167	23	W14575	Streptococcus pneumoniae
20	94	8.4	206	23	W14574	Streptococcus pneumoniae

21	92	8.3	166	23	W14568	Streptococcus pneumoniae
22	93	8.3	183	3	P60623	Sequence A encoded by
23	92	8.3	254	2	R10531	Prod. of DNA of pmc07
24	92	8.3	1247	14	R77548	MeK5 protein.
25	92	8.3	1693	9	R51264	HEV strain protein en
26	92	8.3	1693	18	R91813	Hepatitis E virus str
27	91	8.2	129	23	W14573	Streptococcus pneumoniae
28	91	8.2	147	11	R60799	Rape abscission/dens
29	91	8.2	289	8	R52633	Guinea pig PH-30, 30
30	90	8.1	294	13	R78729	Human bone morphogene
31	90	8.1	388	13	R78734	Human bone morphogene
32	90	8.1	610	22	W18201	Platelet glycoprotein
33	90	8.1	610	9	R51116	Platelet glycoprotein
34	90	8.1	610	9	R56664	Mutant platelet glyco
35	90	8.1	610	16	R89436	Mutated platelet glyco
36	89	8.0	269	7	R40801	NF-116 Beta.
37	89	8.0	609	6	R32131	CW Colburn region po
38	89	8.0	858	6	R30477	Human leukocyte
39	89	8.0	858	21	W14013	Drosophila linocteg
40	87	7.8	170	17	W00498	Papillomavirus E2 bin
41	87	7.8	278	17	R94965	Survival motor neuron
42	87	7.8	294	17	R96981	Survival motor neuron
43	87	7.8	1442	13	R79480	Rat type II collagen.
44	87	7.8	1784	16	R94427	Felvy F6a provirus clo
45	87	7.8	1784	1	R05898	Gene product of first

ALIGNMENTS

RESULT	1	W14570 standard; Protein: 183 AA.
ID	W14570;	
AC	28-OCF-1997 (first entry)	
DE	Streptococcus pneumoniae Pspa central region.	
KW	PSPA; pneumococcal surface protein; vaccine; otitis media;	
KM	meningitis; bacteraemia; pneumonia.	
OS	Streptococcus pneumoniae strain B9739.	
PN	MO9709994-A1.	
PD	20-MAR-1997.	
PF	16-SEP-1996; U14819.	
PR	15-SEP-1995; US-529055.	
PI	(UABR-) UAB RES FOUND.	
PI	Brilles DE, Brooks-Walter A, Crain MJ, Hollingshead S;	
PI	McDaniel LS, Swiatlo E, Tart R, Yother J;	
DR	WPI: 97-202002/18.	
PT	Streptococcus pneumoniae surface protein Pspc and truncated Pspa -	
PT	used in vaccines for protecting animals against S.pneumoniae	
PT	Infection	
PS	Example 6; Fig 13; 296pp; English.	
CC	This sequence shows the central portion, including the C-terminus	
CC	of the alpha-helix region and some of the proline-rich region, of	
CC	strain B9739. Comparison of the N-terminal and central regions	
CC	(W14533-57 and W14562-91) of Pspa polypeptides from different	
CC	pneumococcal strains can be used to divide the strains into several	
CC	families based on sequence homologies. Pspa polypeptides, or	
CC	fragments of them, can be used in vaccines to protect animals	
CC	against S. pneumoniae infection and hence for the prevention of	
CC	diseases such as otitis media, meningitis, bacteraemia and pneumonia.	
CC	The sequence of the 3' half of the Pspa alpha-helical region and the	
CC	immediate 5' tip of the coding sequence are likely to be the critical	
CC	sequences for predicting Pspa cross-reactions and vaccine	
CC	composition.	
CC	Sequence 183 AA:	
DB	Query Match	9.1%; Score 101; DB 23; Length 183;
DB	Best Local Similarity	26.7%; Pred. No. 8.30e+00;
DB	Matches	20; Conservative 19; Mismatches 36; Indels 0; Gaps 0;
OY	11	MTILOAPTPAPSTIPGRGSGPELFTFDPLPEPAAAGRPASRGHRRSRVLYPRV 70

DT 06-JUL-1995 (first entry)
 DE Human NMDAR2 receptor subunit clone NMDA21 protein.
 KW N-methyl-D-aspartate; receptor; human; NMDA; cation-selective; channel;
 KM glutamate; hippocampus; rat; pcdNAL; NMDA receptor; antagonist.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 473..474
 FT /note- "deletion of 17 a.a. from NMDAR2C a.a."
 FT sequence between these residues"
 PN WO9424284-A.
 PD 27-OCT-1994.
 PF 20-APR-1994; U04387.
 PR 20-APR-1993; US-052449.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Daggett LP, Ellis SB, Liaw CM, Lu C;
 DR WPI: 94-341863/42.
 DR N-PSDB: Q79399.
 PT Isolated DNA encoding a human N-methyl-D-aspartate receptor
 PT subunit - used as probes in the identification and isolation of
 PT nucleic acids encoding related receptor subunits.
 PS Claim 11: Page 99-104; 156pp; English.
 CC The amino acid sequence of the novel N-methyl-D-aspartate (NMDA) receptor
 CC 2C (NMDAR2C) gene clone NMDA21. This clone covers the a.a. 249-1040 of
 CC the complete NMDAR2C protein (Q79372) with the exception of a 17 a.a.
 CC deletion of a.a. 722-738 of the NMDAR2C sequence. The NMDA receptor
 CC contains two subunits: subunit R1 (Q79370) and subunit R2 selected from
 CC the subunits 2A (Q79375), 2B (Q79377), 2C and 2D (Q79378). The receptor
 CC forms part of a family of NMDA receptors which have cation-selective
 CC channels and bind glutamate and NMDA. The NMDAR2C gene was obtained by
 CC amplifying cDNA derived from human brain tissues with primers
 CC corresponding to sequences in the rat NMDAR2A receptor DNA and using the
 CC resultant fragments as probes to screen a cDNA library derived from human
 CC hippocampal RNA. 4 basic clones were isolated: NMDA21 (Q79399), NMDA22
 CC (Q79400), NMDA24 (Q79401) and NMDA26 (Q79402). The clones are thought to
 CC be splice variants of each other. Based on the sequence of the 4 clones,
 CC a series of variants (Q79403-7) of the NMDAR2C receptor were constructed.
 CC The expression of the genes allows the reconstruction of the NMDA
 CC receptor. The complete receptor can be used to identify compounds which
 CC bind or are antagonistic to the human NMDA receptor.
 SO Sequence 775 AA;

Query Match 8.6%; Score 96; DB 12; Length 775;
 Best Local Similarity 42.5%; Pred. No. 1.78e+01;
 Matches 17; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

Db 664 appspcprgspcpcpdppepsptgwpdgg9a 703
 QY 18 TPAPSTIPGRSGPEIFTFDPLPEPAAPAGPSASRG 57

RESULT 9
 ID R66065 standard; Protein; 1214 AA.
 AC R66065;
 DT 06-JUL-1995 (first entry)
 DE Human NMDAR2 receptor subunit clone NMDAR2C-delta15-delta1 gene.
 KW N-methyl-D-aspartate; receptor; human; NMDA; cation-selective; channel;
 KM glutamate; hippocampus; rat; pcdNAL; NMDA receptor; antagonist.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 591..592
 FT /note- "5 a.a. deletion of NMDAR2C a.a. sequence
 FT between these residues"
 FT Misc_difference 716..717
 FT /note- "17 a.a. deletion of NMDAR2C a.a. sequence
 FT between these residues"
 PN WO9424284-A.
 PD 27-OCT-1994.
 PF 20-APR-1994; U04387.
 PR 20-APR-1993; US-052449.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Daggett LP, Ellis SB, Liaw CM, Lu C;
 DR WPI: 94-341863/42.
 DR N-PSDB: Q79407.

PT Isolated DNA encoding a human N-methyl-D-aspartate receptor
 PT subunit - used as probes in the identification and isolation of
 PT nucleic acids encoding related receptor subunits.
 PS Claim 11: Page 99-104; 156pp; English.
 CC The amino acid sequence of the novel N-methyl-D-aspartate (NMDA) receptor
 CC 2C (NMDAR2C) gene clone NMDAR2C-delta15-delta1. This clone covers the
 CC same sequence as the NMDAR2C sequence but contains deletions of a.a.
 CC 591-5 and 722-38 of the NMDAR2C protein (R66035). A series of variants
 CC (Q79403-7) of the NMDAR2C receptor were constructed, by recombination,
 CC based on the sequence of the 4 basic clones: NMDA21 (Q79399), NMDA22
 CC (Q79400), NMDA24 (Q79401) and NMDA26 (Q79402). The NMDA receptor
 CC contains two subunits: subunit R1 (Q79370) and subunit R2 selected from
 CC the subunits 2A (Q79375), 2B (Q79377), 2C and 2D (Q79378). The receptor
 CC forms part of a family of NMDA receptors which have cation-selective
 CC channels and bind glutamate and NMDA. The expression of the genes allows
 CC the reconstruction of the NMDA receptor. The complete receptor can be
 CC used to identify compounds which bind or are antagonistic to the human
 CC NMDA receptor.
 SO Sequence 1214 AA;

Query Match 8.6%; Score 96; DB 12; Length 1214;
 Best Local Similarity 42.5%; Pred. No. 1.78e+01;
 Matches 17; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

Db 907 appspcprgspcpcpdppepsptgwpdgg9a 946
 QY 18 TPAPSTIPGRSGPEIFTFDPLPEPAAPAGPSASRG 57

RESULT 10
 ID R66063 standard; Protein; 1219 AA.
 AC R66063;
 DT 06-JUL-1995 (first entry)
 DE Human NMDAR2 receptor subunit clone pcMV-26-scat-21-NotI-24 protein.
 KW N-methyl-D-aspartate; receptor; human; NMDA; cation-selective; channel;
 KM glutamate; hippocampus; rat; pcdNAL; NMDA receptor; antagonist.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 721..722
 FT /note- "17 a.a. deletion of NMDAR2C a.a. sequence
 FT between these residues"
 PN WO9424284-A.
 PD 27-OCT-1994.
 PF 20-APR-1994; U04387.
 PR 20-APR-1993; US-052449.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Daggett LP, Ellis SB, Liaw CM, Lu C;
 DR WPI: 94-341863/42.
 DR N-PSDB: Q79405.

PT Isolated DNA encoding a human N-methyl-D-aspartate receptor
 PT subunit - used as probes in the identification and isolation of
 PT nucleic acids encoding related receptor subunits.
 PS Claim 11: Page 99-104; 156pp; English.
 CC The amino acid sequence of the novel N-methyl-D-aspartate (NMDA) receptor
 CC 2C (NMDAR2C) gene clone pcMV-26-scat-21-NotI-24. This clone covers the
 CC same sequence as the NMDAR2C sequence but contains a deletion of a.a.
 CC 722-8 of the NMDAR2C protein (R66035). A series of variants (Q79403-7)
 CC of the NMDAR2C receptor were constructed, by recombination, based on the
 CC sequence of the 4 basic clones: NMDA21 (Q79399), NMDA22 (Q79400), NMDA24
 CC (Q79401) and NMDA26 (Q79402). The NMDA receptor contains two subunits:
 CC subunit R1 (Q79370) and subunit R2 selected from the subunits 2A
 CC (Q79375), 2B (Q79377), 2C and 2D (Q79378). The receptor forms part of a
 CC family of NMDA receptors which have cation-selective channels and bind
 CC glutamate and NMDA. The expression of the genes allows the
 CC reconstruction of the NMDA receptor. The complete receptor can be used to
 CC identify compounds which bind or are antagonistic to the human NMDA
 CC receptor.
 SO Sequence 1219 AA;

Query Match 8.6%; Score 96; DB 12; Length 1219;
 Best Local Similarity 42.5%; Pred. No. 1.78e+01;
 Matches 17; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

Query Match 8.8%; Score 98; DB 9; Length 520;
Best Local Similarity 41.3%; Pred. No. 1.31e+01;
Matches 19; Conservative 9; Mismatches 15; Indels 3; Gaps 3

Dd 452 a1slpapppgyrg-rhdsgp-1ftfklltpeps-ptcdggaag 494
::| |::| | |::| |::| | |::| | |::| |
Oy 12 TILAPTPAPSTIPGPRRGSGPEIFTFDLPLEPAARAGRSASRG 57

RESULT 5
ID R85943 standard; Protein; 520 AA.
AC R85943;
DE 12-FEB-1996 (first entry)
DT Alpha-1B adrenergic receptor.
KW Alpha-1B adrenergic receptor; antagonist;
KM benign prostate hyperplasia.
OS Homo sapiens.
PN W09528157-A1.
PM 26-OCT-1995.
PF 04-APR-1995; U04203.
PI Branchek TA, Chiu G, Forray CC, Gluchowski C, Hartig PR,
PI Werzel JM.
DR WPJ: 95-373623/48.
DN N-RSDS: T03128.
PT Use of selective alpha 1C adrenergic receptor antagonists - for
PT treating benign prostatic hyperplasia or inhibiting contraction of
PT prostate tissue
PS Disclosure; Page 81-83; 108pp; English.
CC Human alpha-1B adrenergic receptor DNA (see T03128) was cloned into
CC the EcoT site of eukaryotic expression vector pCEXV-3. The
CC resulting plasmid was co-transfected with pGCosneo into LM(tk-),
CC CHO and NIH3T3 cells. Cell lines selectively expressing the receptor
CC (R85943) were used to screen cpds. for antagonist activity.
SQ Sequence 520 AA;

Query Match 8.8%; Score 98; DB 14; Length 520;
Best Local Similarity 41.3%; Pred. No. 1.31e+01;
Matches 19; Conservative 9; Mismatches 15; Indels 3; Gaps 3

Oy 452 a1slpapppgyrg-rhdsgp-1ftfklltpeps-ptcdggaag 494
::| |::| | |::| |::| | |::| | |::| |
Dd 12 TILAPTPAPSTIPGPRRGSGPEIFTFDLPLEPAARAGRSASRG 57

RESULT 6
ID W14569 standard; Protein; 190 AA.
AC W14569;
DE 28-OCT-1997 (first entry)
KW Streptococcus pneumoniae Pspa central region.
KM PSPA: pneumococal surface protein; vaccine; otitis media;
KM meningitis; bacteraemia; pneumonia.
OS Streptococcus pneumoniae strain Bg8743.
FH key Location/Qualifiers
FT Misc_difference 21
FT /note= "unidentified amino acid"
FT Misc_difference 24
FT /note= "unidentified amino acid"
FT Misc_difference 95
FT /note= "unidentified amino acid"
FT Misc_difference 97
FT /note= "unidentified amino acid"
FT Misc_difference 186
FT /note= "unidentified amino acid"
FN W0970994-A1.
PD 20-MAR-1997.
PF 16-SEP-1996; U14819.
PR 15-SEP-1995; US-529055.
PA (UABR-) UAB RES FOUND.
PI Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S,
PI McDaniel US, Swatilo E, Tart R, Tother J;

```
DR      WPI: 97020002/18.  
PT      Streptococcus pneumoniae surface protein PspC and truncated Pspa -  
PR      used in vaccines for protecting animals against S.pneumoniae  
PS      Infection  
PT      Example 6; Fig 13; 296pp; English.  
CC      This sequence shows the central portion, including the C-terminus  
CC      of the alpha-helix region and some of the proline-rich region, of  
CC      pneumococcal surface protein A (PSPA) of Streptococcus pneumoniae  
CC      strain B99163. Comparison of the N-terminal and central regions  
CC      (W14533-5' and W14562-91') of PSPA polypeptides from different  
CC      pneumococcal strains can be used to divide the strains into several  
CC      families based on sequence homologies. PSPA polypeptides or  
CC      fragments of them, can be used in vaccines to protect animals  
CC      against S. pneumoniae infection and hence for the prevention of  
CC      diseases such as otitis media, meningitis, bacteraemia and pneumonia.  
CC      The sequence of the 3' half of the Pspa alpha-helical region and the  
CC      immediate 5' tip of the coding sequence are likely to be the critical  
CC      sequences for predicting Pspa cross-reactions and vaccine  
CC      composition.  
SQ      Sequence 190 AA;  
  
Query Match          8.7%; Score 97; DB 23; Length 190;  
Best Local Similarity 31.4%; Pred. No. 1.53e+01;  
Matches       22; Conservative   15; Mismatches    31; Indels     2; Gaps     2  
  
Db      115 AAPAPAPAPAPAPAPAPAP-kpapapkpapapapapkpkapapkpapapap-apkpek 172  
OY      16 APTPAPSTIPPEPRRSGGEIETFDPLDEPAALAPAGRGSASGHRKSRRLVLPRTVARQL 75  
QY      Db      173 paekxpapakx 182  
        |         ||| :  
QY      QY      76 PVEEPPNPAKR 85  
  
RESULT      7  
ID      R85203 standard; Protein; 1337 AA.  
AC      R85203;  
DT      12-FEB-1996 (first entry)  
DE      huDEP-1.  
KW      Density enhanced Type III receptor-like protein tyrosine phosphatase;  
KM      huDEP-1.  
OS      Homo sapiens.  
PN      M09530008-A1.  
PD      09-NOV-1995.  
PF      03-MAY-1995; U05512.  
PR      03-MAY-1994; US-237940.  
RA      (COLD-) COLD SPRING HARBOR LAB.  
P1      Oestman A, Tonks NK;  
DR      WPI: 95-393079/50.  
DR      N-PDSB: T06027.  
PT      New density enhanced protein tyrosine phosphatase - used to develop  
PT      prods. to modify transcription, translation and/or activity of  
PT      tyrosine phosphatase(s).  
PS      Claim 4; Page 34-38; 51pp; English.  
CC      A cDNA clone was obt'd. (see T06027) from a HeLa cell cDNA library  
CC      that encoded a novel density-enhanced type iii receptor-like ptp,  
CC      designated huDEP-1 (R85203). huDEP-1 is useful for the study of ptps  
CC      and for the development of therapeutic or prophylactic cpds. e.g. for  
CC      prevention of abnormal or malignant cell growth.  
SQ      Sequence 1337 AA;  
  
Query Match          8.7%; Score 97; DB 14; Length 1337;  
Best Local Similarity 32.1%; Pred. NO. 1.53e+01;  
Matches       17; Conservative   13; Mismatches    22; Indels     1; Gaps     1;
```


Release 3:0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MPsrch_pp    protein - protein database search, using Smith-Waterman algorithm
Run on:      Fri Apr 17 13:40:06 1998;      MasPar time 10.54 Seconds
Tabular output not generated.                633.960 Million cell updates/sec

```

```
Title: >US-08-799-910-10
Description: (1-156) from US08799910.pep
Perfect Score: 1114
Sequence: 1 MCHRSCHSPMTLTLOAPTA.....EPSDIALDSTFLQDHAAF 156
```

Scoring table: PAM 150
Gap 11

Searched: 195121 seqs, 42852602 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:      plr55
               1:plr1 2:plr2 3:plr3 4:plr4
```

Statistics: Mean 36.831; Variance 125.421; scale 0.294

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	111	99.7	156	2	JC5537	This is a DE line.	8.18e+11
2	752	67.5	153	2	S53363	This is a DE line.	1.10e-74
3	519	10.5	374	2	S45921	This is a DE line.	1.63e-01
4	114	10.2	451	2	B34768	This is a DE line.	3.09e-01
5	110	9.9	228	2	S53504	This is a DE line.	7.13e-01
6	109	9.8	1729	2	MSRQOQAS	This is a DE line.	8.78e+01
7	102	9.2	733	2	UC0332	This is a DE line.	3.67e+00
8	103	9.2	913	2	S20590	This is a DE line.	3.00e+00
9	102	9.2	1106	2	J00405	This is a DE line.	3.67e+00
10	103	9.2	1206	2	S24407	This is a DE line.	3.00e+00
11	103	9.2	1468	2	S11515	This is a DE line.	3.00e+00
12	101	9.1	268	2	B39429	This is a DE line.	4.49e+00
13	101	9.1	417	2	S03441	This is a DE line.	4.49e+00
14	101	9.1	816	2	F31277	This is a DE line.	4.49e+00
15	100	9.0	333	2	S61849	This is a DE line.	5.49e+00
16	100	9.0	439	2	S51939	This is a DE line.	5.49e+00
17	100	9.0	1492	2	A40333	This is a DE line.	5.49e+00
18	99	8.9	272	2	A47021	This is a DE line.	6.70e+00
19	98	8.8	272	2	S28013	This is a DE line.	8.17e+00
20	98	8.8	351	2	S50754	This is a DE line.	8.17e+00
21	98	8.8	390	1	O0BE77	This is a DE line.	8.17e+00
22	98	8.8	475	2	A51351	This is a DE line.	8.17e+00
23	98	8.8	516	2	JC2332	This is a DE line.	8.17e+00

24	98	8.8	517	2	A4511	This is a DE line.	8.17e+00
25	98	8.8	575	2	S3311	This is a DE line.	8.17e+00
26	98	8.8	1486	2	B3033	This is a DE line.	8.17e+00
27	97	8.7	1239	1	Q0BE10	This is a DE line.	9.95e+00
28	97	8.7	1337	1	I38670	This is a DE line.	9.95e+00
29	97	8.7	1337	2	I52599	This is a DE line.	9.95e+00
30	96	8.6	245	1	W4WL5	This is a DE line.	1.21e+01
31	96	8.6	245	1	W4WL5	This is a DE line.	1.21e+01
32	96	8.6	356	2	E70032	This is a DE line.	1.21e+01
33	95	8.5	234	1	A4DHO	This is a DE line.	1.47e+01
34	95	8.5	316	2	G02424	This is a DE line.	1.47e+01
35	95	8.5	336	2	S55611	This is a DE line.	1.47e+01
36	95	8.5	377	2	A48018	This is a DE line.	1.47e+01
37	95	8.5	449	2	S16748	This is a DE line.	1.47e+01
38	95	8.5	493	2	J70158	This is a DE line.	1.47e+01
39	95	8.5	534	2	S11961	This is a DE line.	1.47e+01
40	95	8.5	643	2	S55610	This is a DE line.	1.47e+01
41	95	8.5	3866	2	B48205	This is a DE line.	1.47e+01
42	95	8.5	3869	2	B48205	This is a DE line.	1.47e+01
43	94	8.4	232	2	J70069	This is a DE line.	1.79e+01
44	94	8.4	389	2	S57000	This is a DE line.	1.79e+01
45	94	8.4	473	2	S50755	This is a DE line.	1.79e+01

ALIGNMENTS

RESULT	1			
ID	JC5537	STANDARD;	PRT;	156 AA.
XX				
AC	xxxxxx			
XX				
DT	01-JAN-1900			
XX				
DE	This is a DE line.			
XX				
CC	A:Experimental source: monocyte			
CC	A:Note: the authors translated the codon CCG for residue 106 as Arg			
SEQUENCE	156 AA; 10914 MM; 127043 CN;			

RESULT	2			
ID	S33363	STANDARD:	PRT:	153 AA.
XX	xxxxxx			
AC				
XX				
DT	01-JAN-1900			
XX				
DE	This is a DE line.			
XX				
CC	A:Cross-references: EMBL:X67644			
CC	C:Genetics:			
CC	A:introns: 70/3			
CC	C:Keywords: transmembrane protein			
CC	SEQUENCE 153 AA; 16875 MW; 122478 CN;			
SO				
Query Match	67.5%;	Score 752;	DB 2;	Length 153;
Best Local Similarity	70.6%;	Pred. NO. 1.10e-74;		

[illegible][illegible]

QY 135 TSEPS 139

RESULT 6
ID MSETGEQASNGEPPDPTTIRVTLTKLDDREATVITIGLQDTIOSLIDGRRMIQSGFORVIAGRVL STANDARD

AC xxxxxx
DT 01-JAN-1900
DE This is a DE line.
XX
SO SEQUENCE 1729 AA; 194986 MW; 14801419 CN;

Query Match 9.8%; Score 109; DB 2; Length 1729;
Best Local Similarity 23.7%; Pred. No. 8.78e-01;
Matches 31; Conservative 30; Mismatches 65; Indels 5; Gaps 5;

Db 774 FLENRGRIPSTSSAPSTSENPPGSPFNS-EDPAIDIRAGRLPLGTRPNRRTVRETVHAPAA 832
QY 13 ILQPTAPSTIPGPRRSGPEITFTDPLPEPAAPACR-PSASRGHKKRSRVLYPRV 71
Db 833 ARAESPNIISLTFTATHTFAPAGF-PLMASSNVSTSGPPGMPIROVVSPTPTTGL 891
QY 72 RR-QLPVEEPNPAKRLFLTLTIVFCQILMAEGVPAPLP-PEDAPNNAASLAPTPVSPVL 129
Db 892 FEEDLSSGSDQ 902
QY 130 EPNLTSEPSD 140

RESULT 7
ID J00532 STANDARD; PRT; 753 AA.

AC xxxxxx
DT 01-JAN-1900
DE This is a DE line.
XX
CC A:Accession: J00532
CC A:Molecule type: mRNA
CC A:Residues: 1-753 <DIN>
CC A:Experimental source: strain Jervis Bay isolate
SO SEQUENCE 753 AA; 82427 MW; 3187938 CN;

Query Match 9.2%; Score 102; DB 2; Length 753;
Best Local Similarity 32.9%; Pred. No. 3.67e+00;
Matches 24; Conservative 14; Mismatches 30; Indels 5; Gaps 5;

Db 271 LPRGNGRPVLPGRKVRGA-PSS-NL-PLPTPOAPPRARERLQNSLHLHAGRONAPRLRP 327
QY 14 LQAPTAPSTIPGPR-RGSGPEITFTDPLPEPAAPACRPSASRGHKKRSRVLYPRV 72
Db 328 RRLR-SDPIDQTR 339
QY 73 ROLPVEEPNPAK 85

RESULT 8
ID S20590 STANDARD; PRT; 913 AA.

AC xxxxxx
DT 01-JAN-1900
DE This is a DE line.
XX
CC TOIG of: s20590 check: 4303 from: 1 to: 913
CC >P1:s20590
CC exo-alpha-stalldase (EC 3.2.1.18) - Actinomyces viscosus

CC C:Species: Actinomyces viscosus

CC C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 09-Sep-1997

CC C:Accession: S20590

CC R:Henningsen, M.; Roggentin, P.; Schauer, R.

CC Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991

CC C:A:Title: Cloning, sequencing and expression of the stalldase gene from Actinomycetes

CC C:A:Reference number: S20590; MUID:92162190

CC C:A:Accession: S20590

CC C:A:Status: preliminary

CC C:A:Molecule type: DNA

CC C:A:Residues: 1-913 <HEI>

CC C:A:Cross-references: EMBL:X62276; NID:g39254; PID:g39255

CC C:Keywords: glycosidase; hydrolase

SO SEQUENCE 913 AA; 96216 MW; 3970532 CN;

Query Match 9.2%; Score 103; DB 2; Length 913;
Best Local Similarity 29.9%; Pred. No. 3.00e+00;
Matches 20; Conservative 14; Mismatches 31; Indels 2; Gaps 2;

Db 816 SPASRNAAPTPKGMPDEID-RPSDGTMAOPTGAP-ARRVDRRRRRPAGCLARDQ 873
QY 16 APTAPSTIPGPRRSGPEITFTDPLPEPAAPACRPSASRGHKKRSRVLYPRV 75
Db 874 RAADPGP 880
QY 76 PVEEPNP 82

RESULT 9
ID J00405 STANDARD; PRT; 1106 AA.

AC xxxxxx
DT 01-JAN-1900
DE This is a DE line.
XX
CC A:Accession: J00405
CC A:Molecule type: DNA
CC A:Residues: 1-1106 <SHI>
CC A:Cross-references: EMBL:X15867
CC C:A:Note: all the codons between two in-frame stop codons are translated; the t
CC A:Note: the gene encoding this protein overlaps uvra gene
SO SEQUENCE 1106 AA; 119484 MW; 4705861 CN;

Query Match 9.2%; Score 102; DB 2; Length 1106;
Best Local Similarity 29.9%; Pred. No. 3.67e+00;
Matches 20; Conservative 16; Mismatches 29; Indels 2; Gaps 2;

Db 569 PGAPGVAGRRPRVPOPRAGRHPLRRGPA-HPPGHTDRLRAGRPLRBAVHRPAPA 627
QY 19 PAPTIPGPRRSGPEITFTDPLPEPAAPACRPSASRGHKKRSRVLYPR-VVROLPV 77
Db 628 GQPPRHR 634
QY 78 EEPNPAK 84

RESULT 10
ID S24407 STANDARD; PRT; 1206 AA.

AC xxxxxx
DT 01-JAN-1900
DE This is a DE line.
XX
CC TOIG of: s24407 check: 501 from: 1 to: 1206
CC >P1:S24407
CC formin isoform IV - mouse
CC C:Species: Mus musculus (house mouse)
CC C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997


```

Matches      22; Conservative      10; Mismatches      25; Indels      3; Gaps      3;

Db      142 CAQTVSLAAQAQPPSPSPPEPPRGSGPPSLAP-GPVAREKGAGKRGPPRGSPSPYRQRER 200
      1: |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy      7 CHPTWTLQAAPT-PAPSTIPGPRGSG-GPEITFDPLLEPPAALAPAGRRSASGHHKRRSR 64

RESULT      13
ID      E30341      STANDARD:      PRT;      417 AA.
AC      xxxxxx
XX      01-JAN-1900
DE      This is a DE line.
XX      TOIG of: e30341      check: 5600      from: 1      to: 417
CC      >P1,E30341
CC      alpha-1-adrenergic receptor - dog (fragment)
CC      C:Species: Canis lupus familiaris (dog)
CC      C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 08-Sep-1997
CC      C:Accession: E30341
CC      R:Libert, F.; Parmentier, M.; Lefort, A.; Dinsart, C.; Van Sande, J.; Maenhaut
CC      Science 244, 568-572, 1989
CC      A:Title: Selective amplification and cloning of four new members of the G protei
CC      A:Reference number: A30341; MUID: 89242119
CC      A:Accession: E30341
CC      A:Status: nucleic acid sequence not shown
CC      A:Molecule type: mRNA
CC      A:Residues: 1-417 <L15>
CC      A:Cross-References: GB:X14050; NID:g902; PID:g833784
CC      C:Superfamily: vertebrate rhodopsin
CC      C:Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein
CC      SEQUENCE      417 AA; 45891 MW; 865241 CN;

Query Match      9.1%; Score 101; DB 2; Length 417;
Best Local Similarity 43.6%; Pred. No. 4,49e+00;
Matches      17; Conservative      10; Mismatches      10; Indels      2; Gaps      2;

Db      353 ALLSLAPAPQPGRRG-RRDSGP-LTFRLRLAERGSPPAG 389
      ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy      12 TILQAPPAPTIPGPRGSGPEITFDPLLEPPAALAPAG 50

RESULT      14
ID      F31277      STANDARD:      PRT;      816 AA.
AC      xxxxxx
XX      01-JAN-1900
DE      This is a DE line.
XX      TOIG of: f31277      check: 4471      from: 1      to: 816
CC      >P1,f31277
CC      regulatory protein qa-1F - Neurospora crassa
CC      N:Alternate names: QA activator
CC      C:Species: Neurospora crassa
CC      C>Date: 26-Apr-1989 #sequence_revision 26-Apr-1989 #text_change 12-Sep-1997
CC      C:Accession: S04256; f31277
CC      R:Geever, R.F.; Huilet, L.; Baum, J.A.; Tyler, B.M.; Patel, V.B.; Rutledge, B.J
CC      J. Mol. Biol. 207, 15-34, 1989
CC      A:Title: DNA sequence, organization and regulation of the qa gene cluster of N
CC      A:Reference number: S04250; MUID: 89293848
CC      A:Accession: S04256
CC      A:Molecule type: DNA
CC      A:Residues: 1-816 <GE2>
CC      A:Cross-references: EMBL:X14603; NID:g3060; PID:g3068
CC      C:Genetics:
CC      A:Gene: qa-1F
CC      C:Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear

```

MUSMUSEUM
(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

Musrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 17 13:39:01 1998; Maspar time 12.21 Seconds

Tabular output not generated. 538.123 Million cell updates/sec

Title: >US-08-799-910-10

Description: (1-156) from US08799910.pep

Sequence: 1 MCHSRCHPTMTILOAPRA.....EPSDYALDSTFLOQHPAA 156

Scoring table: PAM 150

Gap 11

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_organelle 7:sp_phage 8:sp_plant
9:sp_bacteria 10:sp_protist 11:sp_virus 12:sp_vertebrate
13:sp_unclassified

Statistics: Mean 42.617; Variance 98.163; scale 0.434

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	114	100.0	156	2	PRO1 PROTEIN (DIF-2 PR	8.46e-174
2	114	100.0	351	11	HOMOLOGUE OF RETROVIRA	2.07e-02
3	112	10.1	115	9	PREDICTED ORF.	3.65e-02
4	112	10.1	285	1	CELL WALL PROTEIN PREC	3.65e-02
5	111	10.0	3247	11	UL36.	4.84e-02
6	110	9.9	225	11	GENOME, PARTIAL SEQUE	6.41e-02
7	110	9.9	228	8	PROLINE RICH PROTEIN P	6.41e-02
8	108	9.7	452	8	MYROSINASE-BINDING PRO	1.12e-01
9	102	9.2	418	3	SIMILAR TO CUTICULAR C	5.74e-01
10	102	9.2	568	3	ARTICULIN P60.	5.74e-01
11	102	9.2	640	11	VACCINIA VIRUS GENE F1	5.74e-01
12	102	9.2	680	11	MC0181.	5.74e-01
13	103	9.2	913	9	SIALIDASE (EC 3.2.1.18	4.39e-01
14	103	9.2	1711	9	ENDOGLUCANASE A (EC 3.	4.39e-01
15	101	9.1	1426	8	CUT-LIKE 2 (CUX-2).	7.51e-01
16	100	9.0	185	8	PUTATIVE PROLINE-RICH	9.79e-01
17	100	9.0	333	9	HRPV.	9.79e-01
18	100	9.0	439	8	CHITINASE PRECURSOR.	9.79e-01
19	100	9.0	1048	10	CTD-BINDING SR-LIKE PR	9.79e-01
20	100	9.0	1491	12	ALPHA-1 TYPE II' COLLA	9.79e-01

RESULT ID	1	PRELIMINARY:	PRT:	156 AA.
AC	092691;			
DT	01-FEB-1997 (TREMBLER, 02, CREATED)			
DT	01-FEB-1997 (TREMBLER, 02, LAST SEQUENCE UPDATE)			
DT	01-JAN-1998 (TREMBLER, 05, LAST ANNOTATION UPDATE)			
DE	PRO1 PROTEIN (DIF-2 PROTEIN).			
GN	PRO1 OR DIF-2.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; PRIMATES.			
RN	(1)			
RA	SEQUENCE FROM N.A.			
RA	TRAUZOLD A.;			
RL	SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.			
RP	[2]			
RA	SEQUENCE FROM N.A.			
RA	PIETZSCH A., BUECHLER C., ASLANIDIS C., SCHWITZ G.;			
RA	BIOCHEM. BIOPHYS. RES. COMMUN. 235:4-9(1997).			
RP	[3]			
RA	SEQUENCE FROM N.A.			
RA	KONDRATYEV A.D., CHUNG K.N., JUNG M.O.;			
RA	CANCER RES. 56:1498-1502(1996).			
DR	EMBL: Y1451; E33102; -			
DR	EMBL: X96438; E350480; -			
SQ	SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;			

ALIGNMENTS

Query Match	100.0%	Score 114:	DB 2:	Length 156;
Best Local Similarity 100.0%;				
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Db	1	MCHSRCHPTMTILOAPRASTIPGRSGEITFPDLEPAAPAGRSASGRHK 60		
Oy	1	MCHSRCHPTMTILOAPRASTIPGRSGEITFPDLEPAAPAGRSASGRHK 60		
Db	61	RSRRVLYPRVVRQQLVEEENPAKRLFLTLITVFCQIIAEEGVAPLPEDAPVNASL 120		
Oy	61	RSRRVLYPRVVRQQLVEEENPAKRLFLTLITVFCQIIAEEGVAPLPEDAPVNASL 120		
Db	121	APTPVSVLEPENTLSEPSDYALDSTFLOQHPAA 156		
Oy	121	APTPVSVLEPENTLSEPSDYALDSTFLOQHPAA 156		


```

FT CHAIN      8      228      PROLINE RICH PROTEIN.
SQ SEQUENCE   228 AA;  22480 MM;  ZA2B03B9 CRC32;.

Query Match          9.9%; Score 110; DB 8; Length 228;
Best Local Similarity 24.0%; Pred. No. 6,41e-02;
Matches    30; Conservative    36; Mismatches    56; Indels     3; Gaps     3;

Db      154 SSPA 158
Qy      135 TSER 139

RESULT      8      PRELIMINARY; PRT; 552 AA.
ID 096343
AC 096343;
DT 01-FEB-1997 (TREMBLREL_02, CREATED)
DT 01-FEB-1997 (TREMBLREL_02, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL_02, LAST ANNOTATION UPDATE)
DE MYROSINASE-BINDING PROTEIN RELATED PROTEIN (FRAGMENT).
OS BRASSICA NAPUS (RAPE).
OC EUKARYOTA; PLANTIA; EMBRYPHNTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC CAPRARIACE; CRUCIFERAEE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-20516 OF SVALOFS KARAT;
RA TAIPALENSU J., FALK A., EK B., RASK L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U59446; GI655830; -.
FT NON_TER 1 1
SQ SEQUENCE 552 AA; 59270 MM; 6861EF25 CRC32;

Query Match          9.7%; Score 108; DB 8; Length 552;
Best Local Similarity 43.5%; Pred. No. 1,12e-01;
Matches    20; Conservative    11; Mismatches    12; Indels     3; Gaps     3;

Db      295 LRTPAPASPAPGAPAPAPAGSHP-APAPAPAPAGGPPAPAPAG 339
Qy      14 LQAPTAPASTIRPGRGSGPEIFTFDLPPEAPAAPA-G-RPSASRG 57

RESULT      9      PRELIMINARY; PRT; 418 AA.
ID 001662
AC 001662;
DT 01-JUL-1997 (TREMBLREL_04, CREATED)
DT 01-JUL-1997 (TREMBLREL_04, LAST SEQUENCE UPDATE)
DT 01-JUL-1997 (TREMBLREL_04, LAST ANNOTATION UPDATE)
DE SIMILAR TO CUTICULAR COLLAGEN.
GN 128F2.6.
OS CAENORHABDITIS ELEGAN.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WILSON R., AINSICOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA MEDLIN: 94150718.
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,
RA COUTSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA LATREILLE P., LIGHTNING J., SHLOD C., MCMURRAY A., MORIMORE B.,
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFEEN L., ROOPER A.,
RA SAUNDERS D., SHOWNKNEEN R., SMALDON R., SMITH A., SONNHAMMER E.,
RA SPADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.,
```

```

RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MADSEN C., FRONICK B.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF000198; G2047346; -
SQ SEQUENCE 418 AA; 38279 MW; 2587D236 CRC32;

Query Match
Best Local Similarity 37.9%; Pred. No. 5.74e-01;
Matches 22; Conservative 17; Mismatches 15; Indels 4; Gaps 3;

DB 363 AAGGAAPPAPAPAPAPAPAPAPAPAPAPAPAGGSGTGRKKVRVRL--RIVR 417
OY 16 APPAPSTIPGRSGPEIFETDPLPEPAAPAGRPSASRG-HRKRSRVLYPRVVR 72

RESULT 10
ID 027212 PRELIMINARY; PRT; 568 AA.
AC 027212;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE ARTICULIN P60.
OS PSEUDOMICROTHORAX DUBIUS.
OC EUKARYOTA; MITOCHONDRIAL EUKARYOTES; ALVEOLATA; CILIOPHORA; CILIATA;
OC NASSULIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-N5B;
RX MEDLINE: 96032269.
RA HUTTENLAUCH I., GEISLER N., PLESSMANN U., PECK R.K., WEBER K.,
RA STRICK R.;
RL J. CELL BIOL. 130:1401-1412(1995).
DR EMBL: L41557; G767685; -
SQ SEQUENCE 568 AA; 61110 MW; CA5BB376 CRC32;

Query Match
Best Local Similarity 25.0%; Pred. No. 5.74e-01;
Matches 30; Conservative 24; Mismatches 62; Indels 4; Gaps 4;

DB 332 VNPVNVPIEVPVVDVVPVFPOLN-IDVVPVVPVAPVPERIIOQPILEQPLVEQ 390
OY 14 LQAPTPAPSTIPGRSGPEIFETDPLPEPAAPAGRPSASRGHRKRSRVLYPRVVR 73
DB 391 HVPDPVPAQEVIVQOPAVPOPYTVOQEVPIPHVPVPOPYAVP-OPVPTPYAVP 449
OY 74 QLVPEENPAKRLFLTLITVFCOILMAEGVAPLP-PEDAPNASTLAPTPV-SPVLEP 131

RESULT 11
ID 084171 PRELIMINARY; PRT; 640 AA.
AC 084171;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE VACCINIA VIRUS GENE F12L HOMOLOG.
OS ORF VIRUS.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;
OC PARAPOXVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-N22;
RA MERCER A.A.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA SULLIVAN J.T., FLEMING S.B., ROBINSON A.J., MERCER A.A.;
RL VIRUS GENES 11:21-29(1995).
DR EMBL: U34774; G1002997; -
DR EMBL: S82833; G1754760; -
SQ SEQUENCE 640 AA; 70097 MW; 0E08D7DD CRC32;

Query Match
Best Local Similarity 35.7%; Pred. No. 5.74e-01;
Matches 20; Conservative 9; Mismatches 24; Indels 3; Gaps 3;

DB 436 RCFRLAYPREVASQSLPYGKPRMYSKLTGGLALALCEVTSASVKTPLVFPD 491
OY 61 RSRVLYP-RVVRQL-PVEENPAKRLFLTLITVFCOILMAEGV-VPAPLPED 113

RESULT 12
ID 098187 PRELIMINARY; PRT; 680 AA.
AC 098187;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE MC018L.
GN MC018L.
OS MOLUSCUM CONTAGIOSUM VIRUS SUBTYPE 1 (MCV1).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;
OC MOLUSCIPOXVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA SENKEVICH T.G., BUGERT J.J., SISLER J.R., KOONIN E.V., DARAI G.,
RA MOSS B.;
RL SCIENCE 273:813-816(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA SENKEVICH T.G., BUGERT J.J., SISLER J.R., KOONIN E.V., DARAI G.,
RA MOSS B.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U60315; G1491961; -
SQ SEQUENCE 680 AA; 72398 MW; 43AB9271 CRC32;

Query Match
Best Local Similarity 36.0%; Pred. No. 5.74e-01;
Matches 27; Conservative 19; Mismatches 23; Indels 6; Gaps 5;

DB 217 ACAPASAPAPASAPAPASAPASAPASAPASAPASAPARTHKRGTPLRPR 276
OY 6 SCHPTMTIQATPAPSTIPGRSGPEIF-TFDPLPEPAAPAG--RPSASGCHKR 61
DB 277 SKRVV-ARAVR-QLP 289
OY 62 SRVLYPRVVRQLP 76

RESULT 13
ID 059164 PRELIMINARY; PRT; 913 AA.
AC 059164;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE STALIDASE (EC 3.2.1.18) (EXO-ALPHA-STALIDASE) (NEURAMINIDASE)
DE (N-ACYLNEURAMINATE GLYCOHYDROLASE) (ALPHA-NEURAMINIDASE).
GN NAN.
OS ACTINOMYCETES VISCOSUS.
OC PROKARYOTA; FIRMICUTES; IRREGULAR ASPEROGENOUS RODS; CORYNEFORM GROUP.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM43798;
RA HENNINGSSEN M.;
RL SUBMITTED (SEP-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM43798;
RX MEDLINE: 92162190.
RA HENNINGSSEN M., ROGGENTIN P., SCHAUER E.R.;

```


Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1.553
/organism="Homo sapiens"
/clone="260175"

MRNA
BASE COUNT 153 a 165 c 106 g 127 t 2 others
ORIGIN

Query Match 40.3%; Score 495; DB 12; Length 553;
Best Local Similarity 96.4%; Pred. No. 0.00e+00;
Matches 528; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

Db 5 GACAGTAATCAATTTATTTGTTGTCACAGAACTACTAGGAGATGTCAGTGCCTC 64
|||||
CP 1204 GACAGTAATCAATTTATTTGTTGTCACAGAACTACTAGGAGATGTCAGTGCCTC 1145
Db 65 CGTACAGAGCCACCAACCCCAACCCCTCTACCTCGACGCCACCTTAAGGCGACTTCAAG 124
|||||
1144 CGTACAGAGCCACCAACCCCAACCCCTCTACCTCGACGCCACCTTAAGGCGACTTCAAG 1085
125 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGATGCGCGGAGTCTCAGACAGTGA 184
|||||
CP 1084 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGATGCGCGGAGTCTCAGACAGTGA 1025
Db 185 CAGACGAGTGTGAGATCTGAGAGATCAGTACACCTCTTAACCTTACAGCCACACAG 244
|||||
CP 1024 CAGACGAGTGTGAGATCTGAGAGATCAGTACACCTCTTAACCTTACAGCCACACAG 965
Db 245 ACTTCATCCCAAGCCGAGACTCTCTCCCAACCGAGTCTCTCCATTTCTTCTCTACT 304
CP 964 ACTTCATCCCAAGCCGAGACTCTCTCCCAACCGAGTCTCTCCATTTCTTCTCTACT 905
Db 305 TGGCGAGTCCAGAGTCTCTCTCTCCAGAGTCCCAAGCTCAATTAATACCAAG 364
CP 904 TGGCGAGTCCAGAGTCTCTCTCTCCAGAGTCCCAAGCTCAATTAATACCAAG 845
Db 365 ACCGTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAAATTAATATAC 424
CP 844 ACCGTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAAATTAATATAC 785
Db 425 GTACACATCTCCATCAGGAGGAGAGTACATTAATATTAATTAAGGAGCAAT 484
CP 784 GTACACATCTCCATCAGGAGGAGAGTACATTAATATTAATTAAGGAGCAAT 725
Db 485 AAG--GAATTAATTAAGAGCGCTTTCCTTCCAGCGGGGCTGAGCCAGCTGGGCTGT 542
CP 724 AATTAAGAAATTAATTAAGAGCGCTTTCCTTCCAGCGGGGCTGAGCCAGCTGGGCTGT 666
543 GCCTCGGT 550
|||||
CP 665 GCCTCGGT 658

RESULT 2
LOCUS M60982 459 bp mRNA EST 07-JUN-1996
DEFINITION z698109.s1 Pancreatic Islet Homo sapiens cDNA clone 339233 3'
ACCESSION M60982
NID g1367741
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 459)
Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,
Pearson, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasaks, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE MASHU-MERCK EST Project
JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK
Mashu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: EST primer
High quality sequence stop: 226.
Location/Qualifiers
1.459

FEATURES

source
/organism="Homo sapiens"
/note="Organ: pancreas; Vector: pBluescript SK-; Site-1:
ECORI; Site-2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993)
Takeda et al. Cloned unidirectionally. Primer: Oligo dt.
-5' adaptor sequence: 5' GAATTCGGGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'
/db_xref="taxon:9606"
/clone="339233"
/isuse_type="pancreatic islet"
/lab_host="SOBR cells (kanamycin resistant)"
complement(<1..>459)
BASE COUNT 121 a 147 c 82 g 108 t 1 others
ORIGIN

Query Match 36.6%; Score 450; DB 14; Length 459;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 454; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 TTTTCTTTTGTGACAGTAATCAATTTATTTGTTGTCACAGAACTACTAGGCGACT 60
|||||
CP 1216 TTTTCTTTTGTGACAGTAATCAATTTATTTGTTGTCACAGAACTACTAGGCGACT 1157
Db 61 CGACATCGCTCGGTACAGCCGACCAACCCCAACCCCTTACCTCGACGCCACCTTAA 120
CP 1156 CGACATCGCTCGGTACAGCCGACCAACCCCAACCCCTTACCTCGACGCCACCTTAA 1097
Db 121 GGGCACTTCAAGAGTGAAGAGATCTCAGGATCTCATTTGATGCGCGCGAAGTC 180
CP 1096 GGGCACTTCAAGAGTGAAGAGATCTCAGGATCTCATTTGATGCGCGCGAAGTC 1037
Db 181 TCACACAGTAGACAGAGGAGTGAAGATCTGAGAGATCAGTACACCTCTAACTTACG 240
CP 1036 TCACACAGTAGACAGAGGAGTGAAGATCTGAGAGATCAGTACACCTCTAACTTACG 977
Db 241 ACCCAGCCAGAGCTTCAATCCAGCGGAGAGTCTCTCCCAACCGAGTCTCTCCCATTT 300
CP 976 ACCCAGCCAGAGCTTCAATCCAGCGGAGAGTCTCTCTCCCAACCGAGTCTCTCCCATTT 917
Db 301 CTTCCTCTACTTTGGCGGAGTTCCAGATCTCGCTTCCAGCAGTCCCAAGAGTCAAT 360
CP 916 CTTCCTCTACTTTGGCGGAGTTCCAGATCTCGCTTCCAGCAGTCCCAAGAGTCAAT 857
Db 361 GAATACCAAGAGAGCTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAA 420
CP 856 GAATACCAAGAGAGCTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAA 797
Db 421 AATACATATTACGTACACATCTTCATCCCTAGAGAGAC 459
CP 796 AATACATATTACGTACACATCTTCATCCCTAGAGAGAC 758

RESULT 3
LOCUS AA047094 470 bp mRNA EST 06-SEP-1996
DEFINITION zK7402.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone
488522 5'
ACCESSION AA047094
NID g1525011
KEYWORDS EST.

SOURCE	ORGANISM
Eukaryoteae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	human.
REFERENCE	1 (bases 1 to 470)
AUTHORS	Hallier,L., Clark,N., Dubnue,T., Elliston,K., Hawkins,M., Holman,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Tzevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE	The Mashu-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	Contact: Wilson RK Mashu-Merck EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. Putive full length read The vector to vector length is 471 Seq primer: -28W13 revz from Amersham High quality sequence stop: 358. Location/Qualifiers 1..470 /organism="Homo sapiens" /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACGGAGAATTCGCGCCGCCCTTTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo." /db_xref="taxon:9606" /clone_lib="Scotres pregnant uterus NBDPU" /sex="female" /dev_stage="adult" /lab_host="DH10B" <1...>470
MRNA	103 a 82 c 141 g 137 t 7 others
COUNT	IN
Query Match	36.5%; Score 448; DB 14; Length 470;
Best Local Similarity	97.9% Pred. No. 0.00e+00;
Matches 456; Conservative	0; Mismatches 9; Indels 1; Gaps 1;
Db	1 ATATTATAGTATTTATGTACGTCCCTCCTTAGTGATGAGNATGTGTAGCATTAATTATT 60
Oy	738 ATTTRATGTATTTTATGTACGTCCCTCCTTAGTGATGAGATGTGTACGTAAATTATT 797
Db	61 TAACCTATGAAGGAGTGTGAGATGTTCCCTGCCTGATAAATGACAGCTCTTGATTTA 120
Oy	798 TAACTATGAAGGAGTGTGAGATGTTCCCTGCCTGATAAATGACAGCTCTTGATTTA 857
Db	121 TTGAGCTTTGTGNACTGGTGGAGACANGACAACCTGNAACTGGCGCAAGTATGAGANA 180
Oy	858 TTGAGCTTTGTGNACTGGTGGAGAACACACCCTGGAACTGGCGCAAAGTATGAGANA 917
Db	181 AATGGGAGGAGCTCGGCTGNGGGAGAGAGCTCCCGGCTGGAGTGTGGGGGGGCTC 240
Oy	918 AATGGGAGGAGCTCGGCTGNGGGAGAGAGCTCCCGGCTGGAGTGTGGGGGGGCTC 977
Db	241 GTAAGTTAGAGAGTATTCATCCTCCAGACATCTCACTCCGTTCTCTACTGTGTAG 300
Oy	978 GTAAGTTAGAGAGTATTCATCCTCCAGACATCTCACTCCGTTCTCTACTGTGTAG 1037
Db	301 ACTTGCGGCGACCATTAAGATAGAGATCCGATGATCTTCCATCTTTGTAAGTCGCT 360

QY 1038 ACTTCGGCGGACCATTTAGGAATAGAGATCCGTAAGATCCCTCCATCTTTTGAAGTCGCT 1097
 Db 361 TTAGGCTGCGTCGAGAGTAGAGAGGCTTGGTGCGGCTGTCCAGCAGCAGCTGTGC 420
 QY 1098 TTAGGCTGCGTCGAGAGTAGAGGCTTGGGGGCTT-GGTGGCGCTGTCCAGCAGCAGCTGTGC 1156
 Db 421 AGATCGCCTAGTATGTTCTGTGTGACACAAATAAAATGATTTACTG 466
 QY 1157 AGATCGCCTAGTATGTTCTGTGTGACACAAATAAAATGATTTACTG 1202
 RESULT 4 427 bp mRNA EST 06-JUN-1997
 LOCUS zx87f03.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 810724
 DEFINITION 3.
 ACCESSION AA457705
 MID g2180425
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 427)
 AUTHORS Hüller, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maris, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterson, R. and Wilson, R.
 TITLE Washu-Merck EST Project 1997
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Wilson RK
 Washu-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAG Consortium (info@lmage.lnl.gov) for further information.
 Seq primer: -41m3 fwd. Et from Amersham
 High quality sequence stop: 374.
 Location/Qualifiers
 1. 427
 /organism="Homo sapiens"
 /note="Organ: ovary; Vector: p773D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer (5 TGTTCACATCTGAGGAGCGGCGCGGCTTTTCTTTTCTTTTCTTTT 3'), reverse-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bernaldo."
 /db_xref="taxon:9606"
 /clone_1ib="810724"
 /clone_1lb="Soares ovary tumor NbHOT"
 /sex="Female"
 /tissue_type="ovarian tumor"
 /lab_host="DH10B (ampicillin resistant)"
 complement(1. ->427)
 /db_xref="GDB:6041404"
 BASE COUNT 118 a 139 c 78 g 92 t
 ORIGIN
 Query Match 34.4%; Score 422; DB 23; Length 427;
 Best Local Similarity 99.8%; Ped. No. 0.00e+00;
 Matches 423; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 4 GACAGTAAACATTTATTTGTTTCACAGAACTACTAGGCATCTGCAGCTGCTC 63
 1204 GACAGTAAACATTTATTTGTTTCACAGAACTACTAGGCATCTGCAGCTGCTC 1145

D	b6	CGTAGACAGCCCAACACCCCACCCTTACTGCGACACCCACCTAAAGGGACATTAC	123
C	p	1144 CGTAGACAGCCCAACACCCCACCCTTACTGCGACACCCACCTAAAGGGACATTAC	1085
D	b	124 AAGATGGAAGAATCTCAGCGATCTCATTCCTTAATGTGCCGCCGAAGTCTCACAGATGA	183
C	p	1084 AAGATGGAAGAATCTCAGCGATCTCATTCCTTAATGTGCCGCCGAAGTCTCACAGATGA	1025
D	b	184 CAGACGAGATTGAATGCTTGAGAGATGCATGTCACCTCTCTAATTTAGACACCACACGAG	243
C	p	1024 CAGACGAGATTGAATGCTTGAGAGATGCATGTCACCTCTCTAATTTAGACACCACACGAG	965
D	b	244 ACTTCATCCAGCCGGGAGCGTCTCTCCCCACCGAGTCTCTCCATTTCTCTCTACTT	303
C	p	964 ACTTCATCCAGCCGGGAGCGTCTCTCCCCACCGAGTCTCTCCATTTCTCTCTACTT	905
D	b	304 TCGCGCAGTTCCAGGTGTCTCTGCTTCACACGATGCCAACAAAGCTCAATAAATACCAAG	363
C	p	904 TCGCGCAGTTCCAGGTGTCTCTGCTTCACACGATGCCAACAAAGCTCAATAAATACCAAG	845
D	b	364 ACCTGCATTTCACAGCGGGGACAATCTCACACCTTGCATTAAGTTAAATATATTAC	423
C	p	844 ACCTGCATTTCACAGCGGGGACAATCTCACACCTTGCATTAAGTTAAATATATTAC	785
D	b	424 GTAC 427	
C	p	784 GTAC 781	
RESULT	5		
LOCUS	AA043722	460 bp	mRNA
DEFINITION	z559bl1.s1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone		
ACCESSION	A4043722		
NID	g1521646		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;		
	Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae;		
	Homo.		
REFERENCE	1 (bases 1 to 460)		
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maiba,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. Insert Length: 1290 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 381. Location/Qualifiers 1..460 /organism="Homo sapiens" /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(5') primer [5' AACTGGAAGATTGCGCGCCGCTTTTTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library		

constructed by M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="487101"
/clone_lib="Soares pregnant uterus NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(<1.>">460)
/db_xref="GDB:j3760815"

mRNA

BASE COUNT 130 a 146 c 83 g 98 t 3 others

ORIGIN

Query Match 34.2% ; Score 420 ; DB 19 ; Length 460;
Best Local Similarity 98.3%; Pred. No. 0.00e+00;
Matches 451; Conservative 0; Mismatches 3; Indels 5; Gaps 5;

	RESULT	LOCUS	DEFINITION	SOURCE ORGANISM	ACCESSION	NID	KEYWORDS	TITLE COMMENT
REFERENCE AUTHORS	6	AA410666	463 bp mRNA EST	Homo sapiens Eukaryote: mitochondrial eukaryotes; Metazoa: Chordata: Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	AA410666	g2069789	human. EST.	1 (bases 1 to 463)
				Kucaba,T., Lacy,M., Le'N., Lennon,G., Maria,M., Martin,J., Moore,B., Schellenberg K., Steptoe,W., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,K. and Wilson,R.				Mashu-Merck Est Project 1997
				Contact: Wilson RK Mashu-Merck Est Project Washington University School of Medicine				

D	b	300	GGTGTGCGGGGCCCTCCTCCAGAGAGACGCCCTTAACGGCGATTCCTGGGCCCCAC	359
O	y	310	GGTGTGCGGGGCCCTCCTCCAGAGAGACGCCCTTAACGGCGATTCCTGGGCCCCAC	368
D	b	360	CCCTGTGTCCCCTCCTCCAGAGACGCCCTTAACGTACTGCAGAGCCCTCGAGACTACT	419
O	y	369	CCCTGTGTCCCCTCCTCCAGAGACGCCCTTAACGTACTGCAGAGCCCTCGAGACTACT	428
D	b	420	GGACCTCAGCACTTTCTCTCCAGAACAC	447
O	y	429	GGACCTCAGCACTTTCTCTCCAGAACAC	456
R	E	S	AA034911	481 bp mRNA EST 10-MAY-1997
L	O	C	zr25c01.s1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone	
D	E	F	471552.3	
A	C	C	AA034911	
N	I	D	g1506874	
K	E	S	EST.	
S	O	S	human.	
"	G	A	Homo sapiens	
T	A	X	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;	
R	E	F	Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae;	
R	E	F	Homo.	
R	E	F	1 (bases 1 to 481)	
A	A	T	Hillier,L., Clark,N., Duboue,T., Elliston,K., Hawkins,M.,	
A	A	T	Holtman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,	
A	A	T	Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,	
A	A	T	Materston,R., Williamson,A., Wohldmann,P. and Wilson,R.	
J	O	R	Washu-Merck EST Project	
C	O	M	Unpublished (1995)	
T	I	T	Contact: Wilson RK	
J	O	R	Washu-Merck EST Project	
C	O	M	Washington University School of Medicine	
J	O	R	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108	
C	O	M	Tel: 314 286 1800	
J	O	R	Fax: 314 286 1810	
C	O	M	Email: est@wustl.edu	
J	O	R	This clone is available royalty-free through LNL ; contact the	
C	O	M	IMAGE Consortium (info@image.llnl.gov) for further information.	
J	O	R	Insert Length: 1245 Std Error: 0.00	
C	O	M	Seq Primer: -40M13 fwd. from Amersham	
J	O	R	High quality sequence stop: 420.	
C	O	M	Location/Qualifiers	
J	O	R	1. .481	
C	O	M	/organism="Homo sapiens"	
J	O	R	/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;	
C	O	M	Site_2: Eco RI; 1st strand cDNA was primed with a Not I -	
J	O	R	oligo(dT) primer [5'	
C	O	M	AACGTGAAGAATTGCGCGCCCTTTTTTTTTTTTTTTT 3'] ,	
J	O	R	double-stranded cDNA was ligated to Eco RI adaptors	
C	O	M	(Pharmacia), digested with Not I and cloned into the Not I	
J	O	R	and Eco RI sites of the modified pT73 vector. Library	
C	O	M	went through one round of normalization. Library	
J	O	R	constructed by M. Fatima Bonaldo."	
C	O	M	/db_xref="taxon:9606"	
J	O	R	/clone="471552"	
C	O	M	/clone_lib="Soares pregnant uterus NBHPU"	
J	O	R	/sex="female"	
C	O	M	/dev_stage="adult"	
J	O	R	/lab_host="DH10B"	
C	O	M	complement(<1..>481)"	
J	O	R	/db_xref="GDB:3757554"	
C	O	M	BASE COUNT 140 a 139 c 85 g 105 t 12 others	
J	O	R	ORIGIN	
C	O	M	Query Match 32.7% Score 402; DB 19; Length 481;	
J	O	R	Best Local Similarity 95.4%; Pred. No. 0.00e+00;	
C	O	M	Matches 459; Conservative 0; Mismatches 13; Indels 9; Gaps 7;	
D	B	1	GACACTAATCATTTTATTTGTGTTCACAGAACTAGGCGATCTCGACAGTCGCTC	60

CP	1204	GAAGAGTAATCATATTTTATTGTTGTGTACAGAAATCTAGTGGGATCTCGAGACGTGCGTC	1145
Db	61	CGTGACAGCCCAACCAACCCCAACCCCTGTACCTGTGACAGCCACCCCTTAAGGCGACTTCAAG	120
CP	1144	CGTGACAGCCCAACCAACCCCAACCCCTGTACCTGTGACAGCCACCCCTTAAGGCGACTTCAAG	1085
Db	121	AAATGTGAAGGATCTCAGGATCTCATTTCCATTAATGGGCGCGGAAGTCTCACAGATGA	180
CP	1084	AAATGTGAAGGATCTCAGGATCTCATTTCCATTAATGGGCGCGGAAGTCTCACAGATGA	1025
Db	181	CAGACGAGTTTATAGATGCTGTGAGATGAGTACAGTACCTCTTAATATTAAGACCAACAGCAG	240
CP	1024	CAGACGAGTTTATAGATGCTGTGAGATGAGTACAGTACCTCTTAATATTAAGACCAACAGCAG	965
Db	241	ACTTCNTCACNCNCAGNAGGAGCGTCTCNCNCANCNGAGTCTTCCCATTTCTTCTCC	300
CP	964	ACTTCNTC---C-CAGCC-GGGAGAGTCTCTCCCAACCGAGTCTCTCCCATTTCTTCTCC	910
Db	301	TACTTTGCCGNGAGTTCACAGTCTGTCTGTCTTCCACAGTGTCCCAAGATCAATAATAC	360
CP	909	TACTTTGCCGNGAGTTCACAGGT-GTCTGTCTTCCACAGTGTCCCAAGATCAATAATAC	851
Db	361	CAGACAGCTGTGATTTTACAGACAGAGGAGACATNTAANAACCTTTGCATTAAGTTAAAT	420
CP	850	CAGACAGCTGTG-CATTTACAGACAGAGG-AAACATCTCA-CACCTTGCAATAGTTAAAT	794
Db	421	AAATATTACGTACACATCTCCATCCACCTAGAGAGAGCTACATTAATTAATAATATTN	480
CP	793	AAATATTACGTACACATCTCCATCCACCTAGAGAGAGCTACATTAATTAATAATATTN	734
Db	481	A 481	
CP	733	A 733	
RESULT	9		
LOCUS	W52269	413 bp	mRNA
DEFINITION	zc4ta10.s1 Soares senescent fibroblasts NBHSF Homo sapiens CDNA		
ACCESSION	W52269		
NID	91349381		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 413)		
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,O., Maita,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisaks,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The Mashu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT			
	Contact: Wilson RK		
	Washu-Merck EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@wustl.wustl.edu		
	This clone is available royalty-free through LNL; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Insert Length: 3171 Std Error: 0.00		
	Seq primer: MOB.REGA+ET.		
FEATURES	Location/Qualifiers		
SOURCE	1. 413		
	/organism="Homo sapiens"		
	/note="Vector: pT73D (Pharmacia) with a modified		
	polylinker V_Type: phagemid; Site_1: Not I; Site_2: Eco		
	R1: TGTTACCAATCTGAAAGTGGAGCGGCCGCAATTTTTTTTTTTTTTTTTT		

